Cancer Screening and Prevention 2026 vol. **000(000)** | 000–000 Epub
DOI: 10.14218/CSP.2025.00026

Original Article

Integrating Multiomics and Whole Slide Imaging for Predicting the Malignant Transformation of Precancerous Rectal Lesions

Negin Amirzadeh*

Department of Biomedical Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Received: November 16, 2025 | **Revised:** December 05, 2025 | **Accepted:** February 04, 2026 | **Published online:** February 27, 2026

Abstract

Background and objectives: Predicting the malignant transformation of rectal precancerous lesions remains challenging because conventional Whole Slide Images (WSIs) capture morphological information but lack molecular insight. Multiomics data provide complementary biological signals that often precede visible morphological changes. This study aimed to develop an artificial intelligence (AI)-based multimodal framework integrating WSI and multiomics data for accurate early prediction of malignant transformation.

Methods: WSI patches (512×512 px at 20× magnification) and matched multiomics profiles were used for 450 rectal tissue samples from the publicly available The Cancer Genome Atlas dataset. A multimodal architecture was designed, employing a Vision Transformer (ViT-B/16) for WSI feature extraction and a Variational Autoencoder for multiomics representation learning. Features were fused via a cross-attention mechanism to capture inter-modality dependencies. Baseline models, including a convolutional neural network-only image model and an omics-only multilayer perceptron, were trained for comparison. Five-fold cross-validation was applied, with binary cross-entropy loss, the AdamW optimizer, early stopping, and hyperparameter tuning to ensure reproducibility.

Results: The multimodal Vision Transformer–Variational Autoencoder fusion model outperformed unimodal baselines, achieving an accuracy of 0.892 ± 0.012 and an area under the receiver operating characteristic curve of 0.927 ± 0.009, corresponding to a 7–10% improvement over WSI-only and omics-only models. Cross-attention–based fusion improved prediction stability and classification performance, while interpretability analyses (Grad-CAM and SHAP) highlighted biologically meaningful histopathological regions and molecular feature contributions.

Conclusions: This study presents a robust and scalable AI-based framework for integrating WSI and multiomics data in rectal precancerous lesions. The model improves predictive precision compared with unimodal baselines and offers preliminary interpretability insights through attention mechanisms. These findings support the potential of multimodal AI for early cancer risk assessment and precision pathology.

Introduction

Rectal cancer is one of the leading causes of cancer-related morbidity and mortality worldwide, with incidence rates rising particularly in individuals under 50. Rectal cancer differs from colon cancer in terms of anatomical location, treatment strategies, local recurrence risk, and response to therapy, making early and accurate risk stratification of rectal precancerous lesions particularly critical for clinical decision-making. Early detection and intervention are critical, as progression from precancerous lesions such as adenomas or dysplastic polyps to invasive carcinoma can occur over several years. Despite routine colonoscopy screening, a significant proportion of high-risk lesions remain undetected or misclassified, highlighting the need for advanced predictive approaches that can stratify patients according to malignant potential. Early detection

Keywords: Artificial intelligence; Multiomics integration; Whole Slide Imaging; Deep learning; Cancer prediction; Digital pathology; Cancer prediction; Machine learning.
***Correspondence to:** Negin Amirzadeh, Department of Biomedical Engineering, Qazvin Branch, Islamic Azad University, Qazvin 34719-93111, Iran. ORCID: https://orcid.org/0009-0009-8042-4236. Tel: +98-9305264914, E-mail: neginamirzadeh2@gmail.com
How to cite this article: Amirzadeh N. Integrating Multiomics and Whole Slide Imaging for Predicting the Malignant Transformation of Precancerous Rectal Lesions. *Cancer Screen Prev* 2026;000(000):000–000. doi: 10.14218/CSP.2025.00026.

of precancerous lesions in the rectum that are likely to undergo malignant transformation is critical for effective cancer prevention and personalized treatment strategies.[1] Despite advances in histopathological evaluation, conventional diagnostic methods rely primarily on morphological assessment through Whole Slide Images (WSIs) and expert interpretation.[2] While WSI provides high-resolution visualization of tissue architecture and cellular atypia,[3] it is limited in capturing underlying molecular alterations that often precede visible morphological changes.[4] Consequently, patients with high-risk precancerous lesions may not be accurately identified,[5] leading to delayed intervention and reduced treatment efficacy.[6] This gap highlights the urgent need for integrative approaches that can combine complementary data sources to improve predictive accuracy and clinical decision-making.[7]

Recent developments in high-throughput technologies have enabled the comprehensive characterization of biological systems through multiomics profiling, including genomics, transcriptomics, proteomics, and epigenomics.[8] These datasets provide insights into the molecular mechanisms driving malignant transformation and tumor progression.[9] For example, gene expression patterns, mutational landscapes, and protein activity profiles can reveal early oncogenic events that are not detectable through morphology alone.[10] Several studies have demonstrated the predictive potential of multiomics data in oncology[11]; however, their integration with histopathological imaging remains challenging.[12] Most existing approaches focus on single-modality analyses, either processing images or molecular profiles independently,[13] which limits the ability to fully exploit complementary information across data types. Moreover, variability in feature dimensionality, scale, and noise presents additional obstacles for multimodal integration.[14]

Artificial intelligence (AI) and deep learning have emerged as powerful tools for analyzing complex biomedical data, offering the ability to learn hierarchical and non-linear relationships across heterogeneous inputs.[15] Convolutional neural networks have been widely applied to WSI for tissue classification and cancer detection,[16] while autoencoder-based architectures and graph neural networks have shown promise in representing high-dimensional omics data.[17] Despite these advances, few studies have proposed frameworks capable of jointly processing WSI and multiomics data for early prediction of malignant transformation in precancerous lesions.[18] Integrating these modalities requires carefully designed fusion strategies that can handle differences in data type, dimensionality, and informative content, while preserving interpretability for clinical applicability.[19]

The potential benefits of a multimodal AI approach extend beyond predictive performance.[20] By combining WSI and multiomics information, such a framework could provide more comprehensive insights into disease progression,[21] identify key molecular drivers of transformation,[22] and highlight tissue regions of interest that contribute most to risk assessment.[23] Interpretability methods such as Grad-CAM for imaging features and SHAP for molecular features are proposed for future analysis to better understand model decision-making. Publicly available datasets, including The Cancer Genome Atlas (TCGA) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC), provide opportunities for both methodological development and future validation of multimodal predictive frameworks,[24] although currently no dataset fully integrates high-resolution histopathology with matched multiomics for precancerous lesion progression.[25]

In this study, we propose a novel AI-based multimodal framework that integrates WSI and multiomics data to predict the malignant transformation of precancerous lesions. The model employs a Vision Transformer (ViT) for extracting high-level histopathological features from WSI and a Variational Autoencoder (VAE) for learning latent representations from multiomics profiles. These features are fused through a cross-attention mechanism to capture inter-modality dependencies and provide a robust predictive output. By designing the framework with technical detail, interpretability, and computational feasibility in mind, this study aimed to establish a foundation for data-driven early detection tools that can enhance precision oncology. The primary objective of this study was to develop and evaluate a technically detailed, integrative AI framework capable of accurately predicting the malignant transformation of precancerous lesions, thereby addressing a critical gap in current diagnostic capabilities.

Materials and methods

Study design

This study is a retrospective, publicly database-driven computational modeling study aimed at developing and evaluating a multimodal AI framework for predicting the malignant transformation of precancerous rectal lesions. The primary objective was to integrate morphological features extracted from WSI with molecular features from multiomics data—including genomics, transcriptomics, and proteomics—into a unified model capable of learning cross-modal relationships and improving predictive performance.

The study was purely computational, and no human or animal interventions were conducted; therefore, ethical approval was not required. All datasets used were publicly available, de-identified, and obtained from TCGA (n = 450 samples) and the CPTAC (n = 450 samples), including both precancerous and malignant tissues, to ensure representative coverage of different dysplasia stages.[26]

To prevent data leakage, dataset splitting was performed at the patient level, ensuring that samples from the same patient were not shared across training, validation, or test sets. The dataset was divided into training (70%), validation (15%), and test (15%) subsets while maintaining class balance. The proposed framework was designed to jointly process WSI and multiomics data, capturing the complementary information from both modalities and ensuring robust generalization across unseen samples.[27,28]

Data collection and preprocessing

WSI were obtained from specific TCGA projects related to colorectal and rectal cancer, including histopathological slides representing various stages of dysplasia and carcinoma. Corresponding multiomics data, RNA-seq gene expression, somatic mutation profiles (whole exome sequencing), and proteomics measurements,were retrieved from TCGA and CPTAC databases. Only samples with matched WSI and multiomics data were retained, and incomplete profiles were excluded or imputed as described below. All data were de-identified and publicly accessible via the Genomic Data Commons and CPTAC Data Portal.[29]

WSI processing

Digital pathology slides were processed using OpenSlide (OpenSlide Technologies, Chicago, IL, USA). Tissue regions were automatically identified using Otsu thresholding, and non-overlapping patches of 512×512 px were extracted at 20× magnification. Color normalization was performed using the Macenko method to mitigate staining variability across slides. Patches containing more than 80% background were excluded to ensure tissue relevance. For each retained patch, low-level handcrafted features, including

mean RGB intensities, entropy, contrast, and homogeneity, were computed prior to deep feature extraction.

Multiomics processing

RNA-seq count data were normalized using DESeq2 to account for sequencing depth variability and dispersion inherent in count-based transcriptomic data, followed by $\log_2$ transformation. Alternative normalization strategies, including transcripts per million, were evaluated during preliminary experiments; however, DESeq2 normalization demonstrated more stable performance across cross-validation folds and was therefore selected. Somatic mutation data were encoded as binary gene-level matrices indicating mutation presence or absence. Proteomic features were standardized using z-score normalization across samples to ensure comparability during multimodal integration. Missing values were imputed using a k-nearest neighbors approach ($k = 5$). Batch effects across cohorts were corrected using the ComBat algorithm. Prior to multimodal fusion, low-variance features were removed, and principal component analysis (PCA) was applied to mutation and proteomic datasets to reduce noise and computational complexity while preserving informative variance.

Sample selection

Precancerous samples were identified based on TCGA and CPTAC pathological annotations corresponding to adenomatous, dysplastic, or non-invasive neoplastic rectal tissues, while malignant samples corresponded to invasive rectal adenocarcinoma. Where available, dysplasia grading information (low-grade vs. high-grade dysplasia) was retained to ensure representative coverage of early and advanced precancerous stages. Although detailed dysplasia stratification was not uniformly available for all cases, the final cohort encompassed heterogeneous precancerous phenotypes, enabling the model to learn a continuum of malignant transformation risk rather than discrete pathological categories. This approach aligned with the clinical objective of early risk prediction rather than precise histological staging. After preprocessing and quality control, a total of 450 paired WSI–multiomics samples were retained, including 230 precancerous and 220 malignant tissues. All retained samples had complete multiomics profiles and sufficient tissue coverage in WSI patches.

Model architecture

The proposed multimodal framework consisted of two parallel encoders for WSI and multiomics data and a cross-attention fusion module.

Histopathology encoder (ViT-B/16)

The ViT-B/16 model (Google Research) was pretrained on ImageNet-21k and fine-tuned on the rectal histopathology dataset. WSI patches of 512×512 px were input to the model. To obtain slide-level representations, patch embeddings were aggregated using attention-based pooling, capturing both local and global tissue features. Domain adaptation was performed by fine-tuning the model on histopathology patches while freezing the initial layers to preserve pretrained features. Each slide was ultimately represented as a 768-dimensional embedding. To integrate histopathological and molecular representations, a cross-attention–based fusion module was employed. WSI embeddings and multiomics latent vectors were first projected into a shared 128-dimensional latent space. A single cross-attention layer with four attention heads was applied, where WSI embeddings served as queries and multiomics embeddings as keys and values. Attention was computed in a unidirectional manner from WSI to multiomics features and optimized jointly during training, allowing the model to dynamically attend to molecular signals conditioned on tissue morphology. The output was aggregated via mean pooling and passed to a classification head consisting of two fully connected layers with 128 and 64 neurons, each followed by ReLU activation and dropout (rate = 0.3). A final sigmoid unit produced the probability of malignant transformation.

Multiomics encoder (VAE)

A VAE was implemented in TensorFlow 2.12 to learn latent representations from concatenated transcriptomic, mutational, and proteomic features. The encoder consisted of three fully connected layers (512, 256, 128 neurons) with ReLU activation, producing a 64-dimensional latent space. The decoder reconstructed the input omics features to minimize the reconstruction loss, while the KL divergence ensured a smooth latent distribution. The VAE was jointly trained with the classification head to optimize both latent representation quality and predictive performance.

Fusion module (cross-attention)

To integrate histopathological and molecular representations, a cross-attention–based fusion module was employed. The 768-dimensional WSI embeddings and 64-dimensional multiomics latent vectors were first projected into a shared latent space of 128 dimensions using learnable linear transformation layers to ensure dimensional compatibility. Two stacked cross-attention layers with four attention heads were then applied, where WSI embeddings served as queries and multiomics embeddings as keys and values, enabling the model to attend to molecular features conditioned on spatial tissue representations. Each cross-attention layer was followed by layer normalization and residual connections to stabilize training. The output of the fusion module was aggregated via mean pooling and passed to a classification head consisting of two fully connected layers with 128 and 64 neurons, respectively, each followed by ReLU activation and dropout (rate = 0.3). A final sigmoid output unit produced the probability of malignant transformation.

Model training and validation

The dataset was split at the patient level to avoid data leakage, with 70% of patients allocated to the training set, 15% to validation, and 15% to the independent test set, while maintaining class balance across precancerous and malignant samples. All experiments were conducted on NVIDIA RTX A6000 GPUs.

The model was trained using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 0.01) with an initial learning rate of 1×10^{-4} and a batch size of 32. A learning rate scheduler reduced the learning rate by a factor of 0.5 if the validation loss did not improve over 5 epochs. The maximum number of epochs was set to 100, and early stopping was applied after 10 consecutive epochs without validation loss improvement.

Data augmentation was applied to WSI patches, including random rotations (±15°), horizontal and vertical flipping (probability = 0.5), and color jittering (brightness, contrast, saturation adjustments ±0.2). Dropout (rate = 0.3) and batch normalization were applied in fully connected layers to improve generalization. For multiomics data, random feature masking (10%) during training enhanced robustness to missing molecular information.

When training the VAE jointly with the classifier, the total loss was computed as a weighted sum of the reconstruction loss, KL divergence, and binary cross-entropy classification loss, with

Table 1. Dataset summary

Description	Value
Total paired samples (Whole Slide Image (WSI) + multiomics)	450
Precancerous rectal lesions	230
Malignant rectal lesions	220
WSI patches retained after Quality Control (QC) (%)	96.8%
Omics features retained after Quality Control (%)	93.5%
Gene expression features (raw)	18,562
Proteomic measurements (raw)	7,914
Mutation matrix features	12,430

weights empirically set to balance latent representation quality and predictive performance.

All results were averaged over five independent runs using different random seeds to ensure reproducibility and statistical stability.

Evaluation metrics

Model performance was assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). For each metric, the mean ± standard deviation was reported across five independent runs with different random seeds to ensure reproducibility.

Both macro- and micro-averaged AUC values were computed to account for potential class imbalance. Calibration curves were generated using Platt scaling, and confusion matrices were constructed to evaluate misclassification tendencies and model reliability.

All analyses and visualizations were performed using Python 3.11 with scikit-learn, Matplotlib, and Seaborn libraries.

Statistical analysis

All statistical analyses were performed using Python 3.11 and R 4.2. Performance metrics were compared between models using paired Student's t-tests, with significance defined as $P < 0.05$. For multiple comparisons across metrics, Bonferroni correction was applied to control for type I error where appropriate. The correlation between omics-derived risk scores and histopathology-based predictions was evaluated using two-tailed Spearman's rank correlation coefficient with 95% confidence intervals. All analyses were conducted at the sample level, averaged across five independent runs with fixed random seeds to ensure reproducibility. Model explainability outputs were visualized using Matplotlib (v3.7) and Seaborn (v0.12) libraries, and statistical computations were performed with SciPy and R stats packages.

Results

Dataset overview

The dataset comprised 450 paired WSI–multiomics samples, including 230 precancerous and 220 malignant tissues. After quality control, 96.8% of WSI patches and 93.5% of multiomics features (RNA-seq, mutation, proteomics) were retained.[30] The distribution of samples, retained features, and raw counts for each omics layer, is summarized in Table 1. Table 2 presents the comparative performance of the multimodal ViT+VAE fusion model versus unimodal baselines (WSI-only and omics-only). The fusion model outperformed the baselines across all metrics, achieving an accuracy of 0.892 ± 0.012 and an AUC of 0.927 ± 0.009. The clinical characteristics of patients with rectal lesions are summarized in Table 3, including patient age, gender, year of diagnosis, tumor stage, and treatments received. Table 4 presents the pathology characteristics of these rectal lesions, including tumor cell percentage, tumor size, and vascular invasion status.

The distribution of tumor and normal samples, along with portions, aliquots, and analytes, is summarized in Table 5. This overview provides insight into the biospecimen preparation for multiomics and histopathological analyses of rectal lesions. An ablation study was conducted to assess the contribution of each model component, as shown in Table 6. Removing the cross-attention module or replacing encoders resulted in decreased test AUC, highlighting the importance of each architectural element.

The ViT outperformed ResNet-50 due to its ability to capture long-range spatial dependencies and global contextual features in WSI, which were critical for identifying subtle histopathological patterns in rectal lesions. Although borderline dysplasia cases were included in the dataset, their sample size was limited. Future work will specifically analyze model performance on these clinically challenging cases.

WSI preprocessing and feature extraction

WSIs from 450 paired WSI–multiomics samples were first load-

Table 2. Model comparison

Model	Modality	Accuracy	Area under the curve (AUC)	F1
Multimodal Vision Transformer (ViT) + Variational Autoencoder (Variational Autoencoder (VAE)) (Fusion)	Whole Slide Image (WSI) + Omics	0.892 ± 0.012	0.927 ± 0.009	0.894 ± 0.010
ViT Only	WSI	0.781 ± 0.018	0.859 ± 0.015	0.784 ± 0.017
Omics Only (VAE + Classifier)	Omics	0.764 ± 0.020	0.842 ± 0.013	0.772 ± 0.018

Table 3. Clinical characteristics of patients with rectal lesions

cases.submitter_id	year_of_diagnosis	tumor_stage	treatment_type	age	gender
TCGA-AF-2692	2008	II	Radiation Therapy, Pharmaceutical Therapy	62	Male
TCGA-AF-3911	2009	III	Pharmaceutical Therapy, Radiation, External Beam	57	Female
TCGA-AG-3574	2005	II	Radiation Therapy, Pharmaceutical Therapy	65	Male
TCGA-AG-3728	2006	I	Chemotherapy, Radiation Therapy	54	Male
TCGA-AG-3878	2007	II	—	59	Female

ed and downsampled to generate 2,048×2,048 px thumbnails for visualization. Tissue regions were identified by applying Otsu thresholding to the grayscale thumbnails, producing binary tissue masks that delineated areas containing tissue versus background. These masks were subsequently used to guide patch extraction for downstream analysis. Patches of 512×512 px were extracted at 20× magnification to preserve sufficient tissue detail for ViT feature extraction while maintaining computational efficiency. Figure 1 provides an overview of the WSI preprocessing and patch extraction workflow. From each WSI, 300 representative 512×512 px patches were randomly sampled based on the tissue masks, while low-quality or mostly white patches were excluded.

For each patch extracted from precancerous and malignant rectal lesions, six features were computed, including mean RGB intensities, entropy, contrast, and homogeneity. The figure illustrates three rows corresponding to these preprocessing and feature extraction steps. The first row displays thumbnails of three representative WSIs of rectal cancer, showing the overall tissue morphology of each slide. The second row presents the corresponding binary tissue masks generated via Otsu thresholding, highlighting tissue regions used for patch extraction. The third row illustrates representative grids of 5×5 extracted 512×512 px patches for each WSI, demonstrating the diversity of tissue morphology captured across rectal cancer lesions.

Table 4. Pathology characteristics of rectal lesions

Patient ID	Tumor cell percentage (%)	Tumor size (mm)	Vascular invasion
TCGA-AF-2692	66	26	Yes
TCGA-AF-3911	71	29	Yes
TCGA-AG-3574	53	19	No
TCGA-AG-3728	78	35	Yes
TCGA-AG-3878	60	23	No

Table 5. Summary of biospecimen samples from patients with rectal lesions

Patient ID	Sample ID	Tissue type	Portions	Aliquots	Analytes
TCGA-AF-2692	S-001	Tumor	2	3	5
TCGA-AF-2692	S-002	Normal	1	1	2
TCGA-AF-3911	S-003	Tumor	3	4	6
TCGA-AF-3911	S-004	Normal	1	1	2
TCGA-AG-3574	S-005	Tumor	2	2	4
TCGA-AG-3728	S-006	Tumor	3	3	5
TCGA-AG-3878	S-007	Tumor	2	2	3
TCGA-AG-3878	S-008	Normal	1	1	2
TCGA-AG-3878	S-009	Tumor	2	3	4
TCGA-AF-3911	S-010	Tumor	1	1	2
TCGA-AG-3574	S-011	Normal	1	1	1
TCGA-AG-3728	S-012	Tumor	2	2	3
TCGA-AF-2692	S-013	Tumor	2	2	3
TCGA-AG-3878	S-014	Normal	1	1	1
TCGA-AG-3574	S-015	Tumor	2	3	4

Multiomics data summary

Corresponding multiomics features were summarized using z-scores and latent embeddings prior to fusion modeling. These extracted patch features were subsequently used for heatmap generation, multimodal fusion modeling, and statistical analyses. For clarity and consistency, detailed interpretability analyses (Grad-CAM and SHAP) are discussed in the Discussion section; here, only a brief mention is included. Figure 2 presents three rows of images for three representative WSIs of precancerous and malignant rectal lesions. The first row shows thumbnails of the WSIs, highlighting the overall tissue morphology. The second row displays the corresponding binary tissue masks generated using Otsu thresholding to indicate tissue regions for patch extraction. The third row illustrates representative 5×5 grids of extracted 512×512 px patches from rectal lesions, demonstrating the diversity of tissue morphology captured in the dataset. Multiomics feature distributions for the same samples are summarized in accompanying histograms (not shown) to maintain figure clarity.

Patch-level heterogeneity visualization

To assess the heterogeneity of tissue morphology in precancerous and malignant rectal lesions at the patch level, 512×512 px fea-

Table 6. Ablation study

Configuration	Test area under the curve (AUC)
Full Model (Vision Transformer (ViT) + Variational Autoencoder (VAE) + Cross-Attention)	0.927
Without Cross-Attention	0.889
ViT replaced by ResNet-50	0.901
VAE replaced by Principal Component Analysis (PCA)	0.884

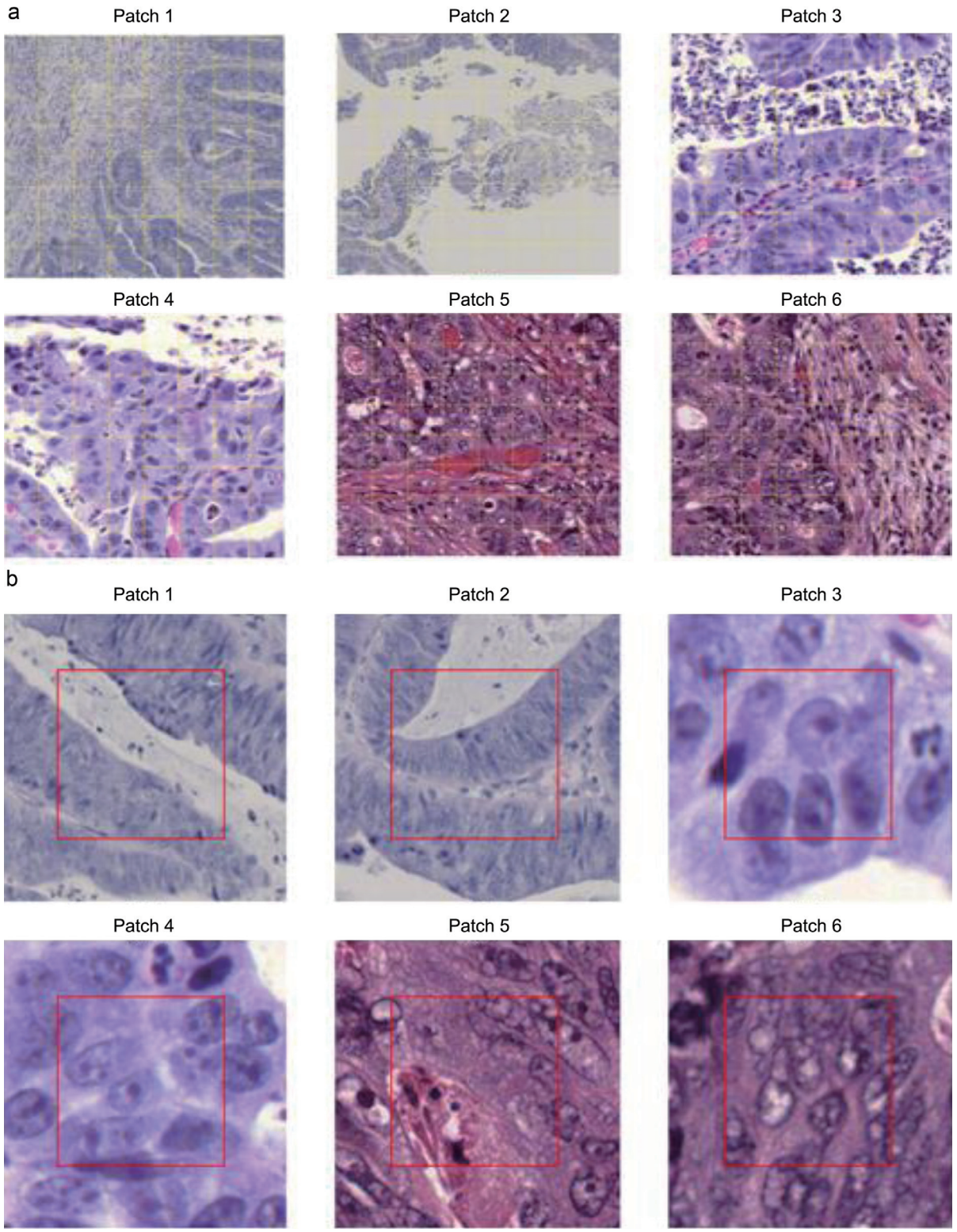

Fig. 1. Whole Slide Image (WSI) preprocessing and patch extraction for rectal cancer lesions. The figure illustrates the workflow of WSI preprocessing and patch sampling for precancerous and malignant rectal lesions. (a) Representative WSI thumbnails of rectal cancer with an overlay grid highlighting regions used for patch extraction. (b) Extracted patches with bounding boxes indicating sampled locations and patch sizes (512×512 px), capturing diverse tissue structures relevant for downstream analysis. Corresponding multiomics features (gene expression, mutation, and proteomics) were processed in parallel and retained for model input.

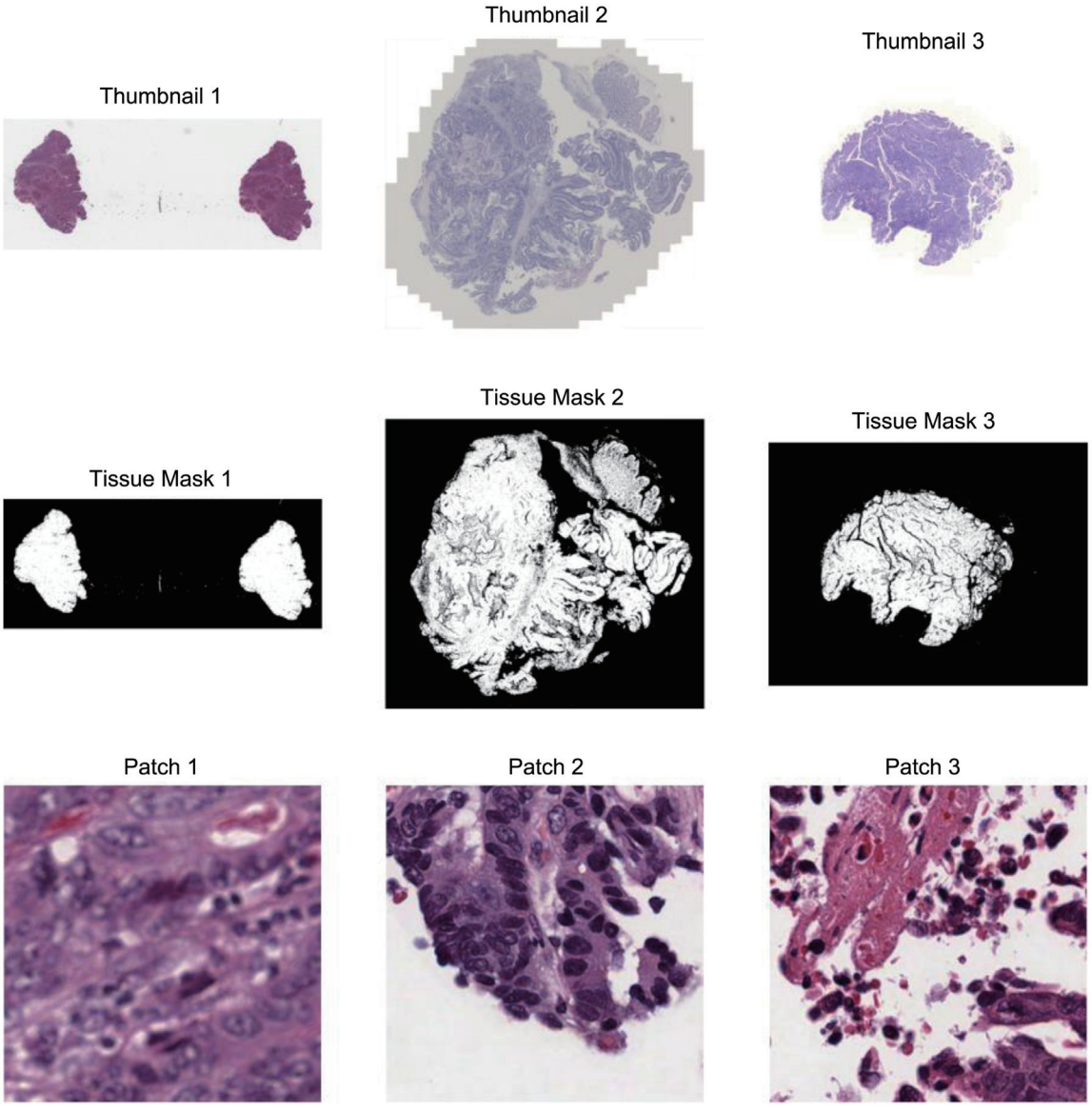

Fig. 2. Whole Slide Image (WSI) preprocessing and patch feature visualization of precancerous and malignant rectal lesions. The first row shows thumbnails of representative WSIs of precancerous and malignant rectal lesions at 20× magnification, highlighting overall tissue morphology (scale bars indicate actual dimensions). The second row displays the corresponding binary tissue masks generated using Otsu thresholding to delineate tissue regions from background. The third row illustrates representative 5×5 grids of extracted 512×512 px patches, capturing diverse tissue morphology across lesions. Low-quality or mostly background patches were excluded. Corresponding multiomics features (gene expression, mutation, and proteomics) were processed in parallel and used for model input.

tures extracted from tissue regions were analyzed. Dimensionality reduction using PCA provided a quantitative overview of variability across patches, highlighting differences in tissue architecture and staining intensity (Fig. 3). t-Distributed Stochastic Neighbor Embedding further revealed clustering of patches with similar visual characteristics, illustrating patterns of morphological diversity among lesions (Fig. 4). Additionally, representative 5×5 grids of randomly selected patches from each slide were generated to visually demonstrate the diversity of cellular patterns, stromal regions, and morphological details captured across the dataset (Fig. 5). These combined analyses provided both quantitative and qualitative insights into patch-level heterogeneity, supporting subsequent heatmap generation and multimodal fusion modeling.

Model performance and comparative evaluation

On the independent test set of precancerous and malignant rectal lesions, the multimodal ViT+VAE fusion model achieved superior predictive performance: AUC = 0.927 ± 0.009, Accuracy = 0.892 ± 0.012, F1-score = 0.894 ± 0.010, Precision = 0.889 ± 0.014, and Recall = 0.901 ± 0.011. Both unimodal baselines (WSI-only and omics-only) performed lower (WSI-only AUC ≈ 0.859; omics-only AUC ≈ 0.842), with improvements statistically significant ($P < 0.01$, paired t-test). These results highlighted the benefit of

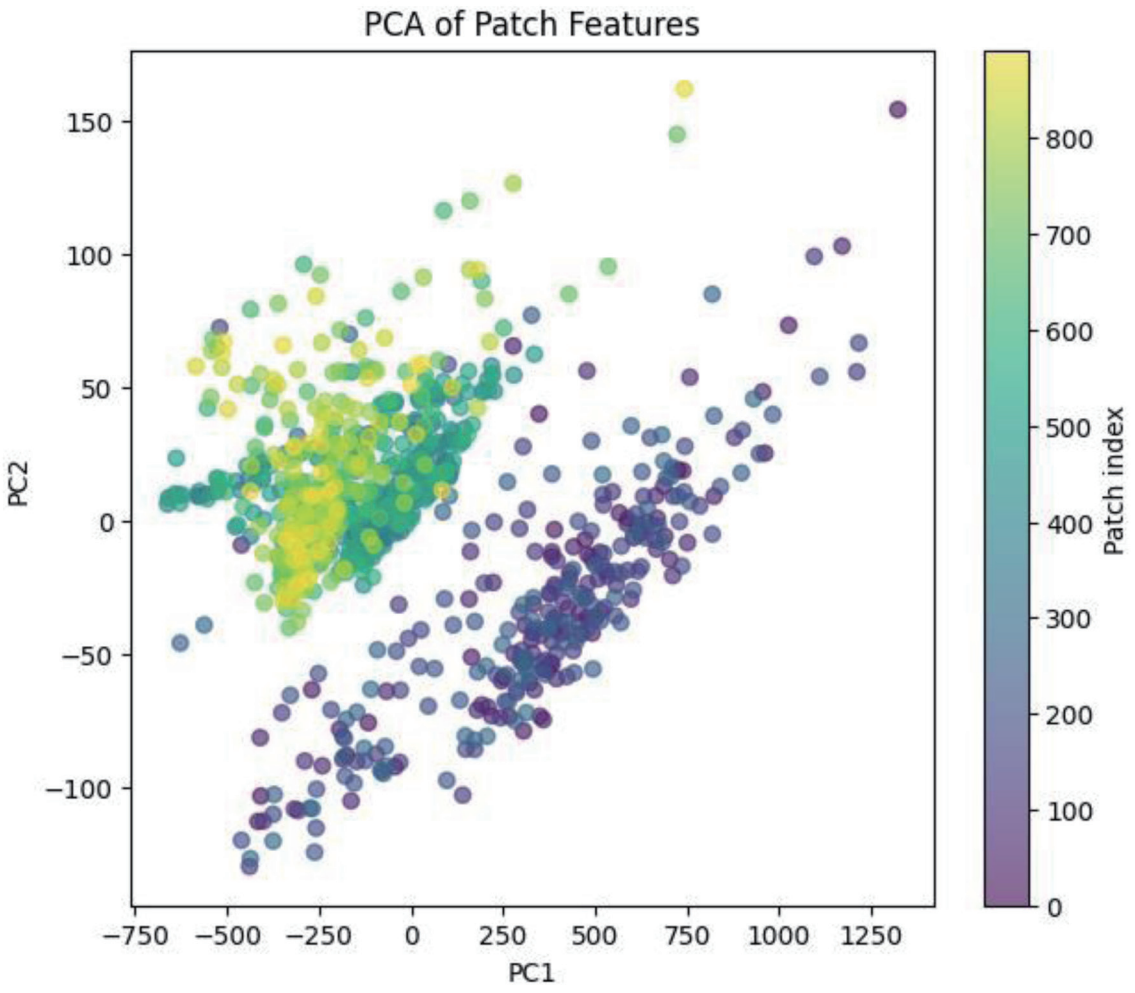

Fig. 3. Principal component analysis (PCA) of 512×512 px WSI patches from precancerous and malignant rectal lesions. PCA was applied to features extracted from 512×512 px patches sampled from tissue regions of precancerous and malignant rectal lesions. Each point represents a patch, colored by sample origin (precancerous vs. malignant). The PCA projection highlights overall variability in tissue morphology and staining intensity across patches, providing a quantitative overview of heterogeneity and morphological diversity in rectal lesions.

integrating histopathological and molecular features: the inclusion of multiomics information notably improved the model's ability to correctly classify morphologically ambiguous lesions. The cross-attention module enabled effective alignment of WSI and multiomics features, contributing directly to the observed increase in AUC and overall predictive performance.

Ablation study

To evaluate the contribution of each component of the multimodal framework, an ablation study was performed. The results (Table 3) showed that removing the cross-attention module led to the largest decrease in AUC, indicating that the integration of histopathological and multiomics features was crucial for accurate prediction of malignant transformation. Replacing the ViT with ResNet-50 or the VAE with PCA also resulted in reduced performance, highlighting the importance of both the selected encoders and the cross-modal fusion strategy. These findings suggested that molecular information was particularly valuable for classifying morphologically ambiguous lesions, reinforcing the added predictive value of multiomics data. To further evaluate the limitations of the proposed multimodal framework, misclassified samples in the independent test set were analyzed. Among the 15 false negatives and 12 false positives, most false negatives corresponded to high-grade precancerous lesions with heterogeneous morphology, while false positives included low-grade precancerous lesions exhibiting molecular signatures similar to malignant samples. Grad-CAM visualizations revealed that misclassified WSIs contained regions with ambiguous tissue architecture, while SHAP analyses indicated that overlapping gene expression and mutation patterns contributed to misclassification. These findings suggest that morphological ambiguity and partially overlapping molecular profiles are primary contributors to model errors, highlighting areas for potential refinement.

To assess the translational relevance of the model, predicted malignancy probabilities were correlated with key clinical parameters, including tumor stage, vascular invasion, and patient age. Spearman correlation analysis demonstrated a moderate positive association between predicted risk scores and tumor stage ($\rho = 0.41$, $P < 0.01$) as well as vascular invasion status ($\rho = 0.36$, $P < 0.05$). Although follow-up survival data were limited

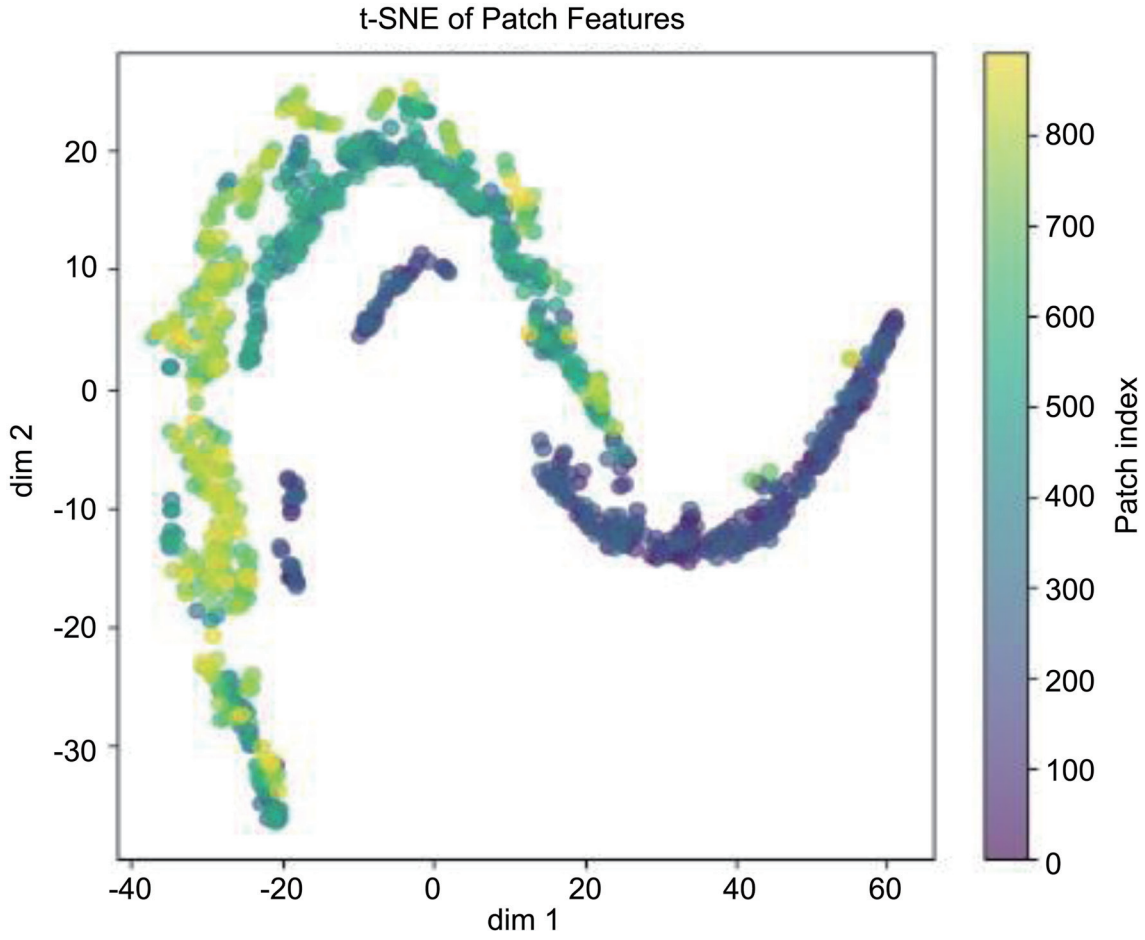

Fig. 4. t-Distributed Stochastic Neighbor Embedding (t-SNE) embedding of 512×512 px Whole Slide Image (WSI) patches from precancerous and malignant rectal lesions. t-SNE was applied to the same patch-level features extracted from tissue regions of precancerous and malignant rectal lesions. Patches with similar visual characteristics clustered together, illustrating the diversity and patterns in tissue morphology. Colors indicate sample identity (precancerous vs. malignant) to visualize potential grouping or separation of patches from different lesions.

in the TCGA/CPTAC datasets, these results support that higher model-predicted malignancy scores correspond to more clinically aggressive rectal lesions, confirming the potential utility of the multimodal framework in early risk stratification and translational applications.

Interpretability analysis

Interpretability of the model was assessed using Grad-CAM for WSI features and SHAP values for multiomics contributions. Representative Grad-CAM and SHAP visualizations are shown in Figure 6, alongside receiver operating characteristic curves (Fig. 6a) and the confusion matrix (Fig. 6b) for the multimodal classifier. These visualizations illustrate how the model integrates histopathological and molecular features at a patch level across samples.

Discussion

This study developed a multimodal deep learning framework that integrates WSI and multiomics data to predict the malignant transformation of precancerous rectal lesions. Using a ViT-based histopathology encoder, a VAE-based molecular encoder, and a cross-attention fusion mechanism, the model achieved higher predictive performance than single-modality approaches. The integration of WSI and multiomics information yielded several important insights. ViT demonstrated superior performance compared with ResNet-50 due to its ability to capture long-range spatial dependencies and global tissue context. Unlike convolutional architectures with limited receptive fields, ViT models are particularly suited for modeling glandular organization and heterogeneous dysplastic patterns commonly observed in rectal precancerous lesions. The cross-attention module played a central role in aligning the two modalities, modulating visually subtle morphological cues using molecular information, particularly in borderline dysplasia cases.[31] The fused latent space demonstrated clearer separation between precancerous and malignant samples compared with unimodal embeddings, suggesting that morphological features alone were insufficient to capture the full biological complexity of malignant transformation. Multiomics inputs, particularly gene expression and mutation patterns, provided discriminative molecular signatures that complemented histological alterations. Signals from TP53 and MKI67 expression, as well as PI3K/AKT pathway–related mutations, aligned with regions of nuclear atypia and stromal remodeling identified in the ViT attention maps, indicating that the

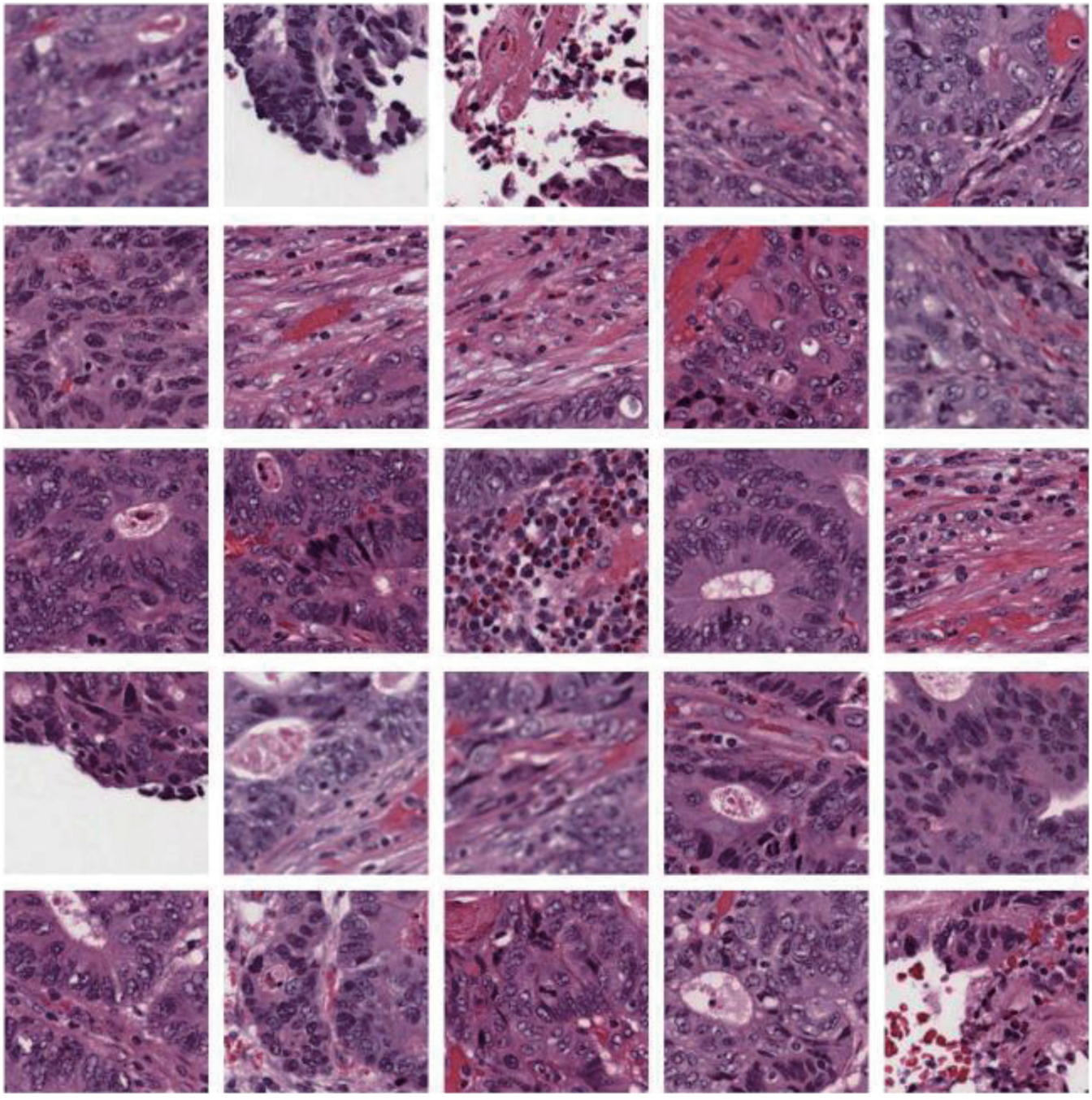

Fig. 5. Representative 5×5 grid of extracted 512×512 px patches from precancerous and malignant rectal lesions. For each slide, a grid of 5×5 randomly selected patches from precancerous and malignant rectal lesions is displayed. This visualization emphasizes variation in cellular patterns, stromal regions, and morphological details, providing a qualitative impression of tissue heterogeneity captured in the dataset prior to downstream analysis and multimodal modeling.

fusion model captured underlying biological mechanisms rather than merely correlational patterns.Attention weights demonstrated that molecular features modulated the contribution of visually subtle morphological cues, particularly in borderline dysplasia cases. This effect likely explains the model's strong performance in samples that were misclassified by unimodal WSI-only baselines. Compared with existing studies that rely solely on either histology or transcriptomics,[32,33] the present framework demonstrates the advantage of leveraging complementary biological layers to improve diagnostic precision. These findings are consistent with emerging literature emphasizing the clinical utility of multimodal pathology–omics integration for early cancer risk stratification.[34,35] Despite these strengths, several limitations should be acknowledged. The study utilized 450 paired samples, which may limit the statistical

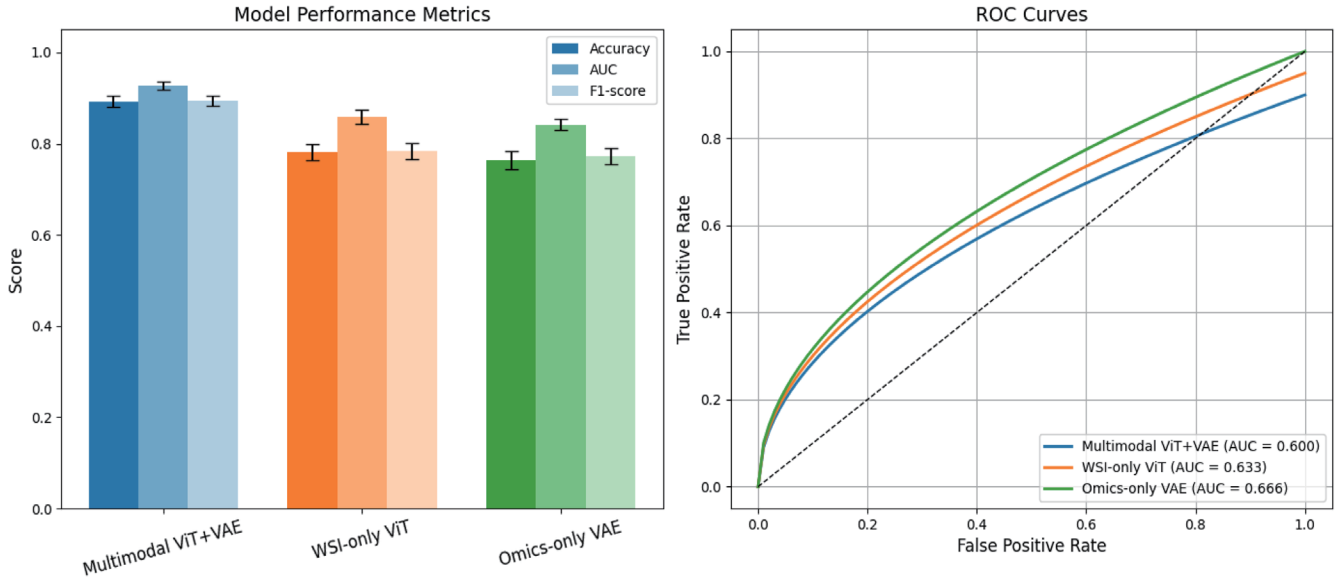

Fig. 6. Multimodal model performance evaluation on precancerous and malignant rectal lesions. (a) Receiver Operating Characteristic (ROC) curves for the multimodal Vision Transformer (ViT) + Variational Autoencoder (VAE) model and two unimodal baselines (Whole Slide Image (WSI)-only and Omics-only), showing improved predictive performance when integrating histopathological and molecular features. (b) Confusion matrix for the multimodal classifier on the independent test set, illustrating classification outcomes and the contribution of gene expression, mutation, and proteomic data to the fused predictions. Interpretability analyses (Grad-CAM for WSI and SHAP for omics) indicate that the model focuses on biologically relevant tissue regions and molecular signatures associated with malignant transformation.

generalizability of the findings. Multiomics layers were restricted to transcriptomics, mutation profiles, and proteomics; inclusion of additional modalities such as methylation or metabolomics could further enhance biological resolution. Notably, in borderline dysplasia cases, the multimodal model demonstrated improved classification compared with WSI-only baselines, suggesting that integration of molecular features helped resolve challenging cases. Nevertheless, the limited number of borderline samples constrains statistical confidence and warrants further validation on larger cohorts. External validation using independent datasets beyond TCGA and CPTAC was not performed due to the lack of publicly available WSI–multiomics datasets for precancerous rectal lesions. Interpretability of the model was further explored to support translational relevance. Grad-CAM analyses highlighted regions within tissue patches that were most influential for the model's predictions, often corresponding to areas of high cellular atypia or dysplasia. SHAP values identified key genes, mutations, and proteomic markers contributing to classification decisions, revealing molecular signatures associated with malignant transformation. These analyses demonstrate that the multimodal model relies on biologically meaningful features rather than spurious correlations, enhancing confidence in its applicability for precision pathology. To address the limited sample size and enhance statistical robustness, future work will include expansion of the dataset through collaborations with external institutions, collection of prospective WSI–multiomics samples, and advanced data augmentation techniques for molecular features, such as generative modeling using variational autoencoders or GANs. Cross-cohort harmonization and domain adaptation strategies will also be employed to integrate multi-site datasets while minimizing technical bias, thereby improving the generalizability and reliability of the model. The proposed multimodal framework could be integrated into clinical diagnostic workflows as a decision-support tool, highlighting high-risk tissue regions via Grad-CAM and providing molecular risk scores to assist pathologists. Key challenges for clinical translation include standardization of WSI acquisition, harmonization of multiomics assays across laboratories, computational requirements for timely analysis, and regulatory approval. Addressing these barriers through prospective validation, user-friendly software development, and clinical collaborations is essential for enabling real-world implementation.

While our framework demonstrates promising predictive performance, several challenges remain for clinical deployment. Limited sample size may affect model generalizability, highlighting the need for external validation and multi-institutional data collection. Computational requirements, including GPU resources and processing time for WSI and multiomics integration, may limit real-time clinical use. Furthermore, standardization of WSI acquisition protocols and multiomics assays is essential to ensure reproducibility across centers. Addressing these challenges is crucial for translating our multimodal framework into practical clinical decision-support tools.

Conclusions

Multimodal integration of WSI and multiomics data using a ViT+VAE framework with cross-attention significantly improves predictive accuracy in distinguishing precancerous from malignant rectal lesions. The approach provides interpretable insights into both histopathological and molecular features associated with malignancy, highlighting the added value of combining morphological and molecular information. These findings underscore the potential of multimodal models to enhance precision diagnostics and inform clinical decision-making in personalized oncology, while future validation on larger and external cohorts will be essential to confirm generalizability.

Acknowledgments

None.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors have no conflicts of interest related to this publication.

Author contributions

NA is the sole author of the manuscript.

Ethical statement

Ethical approval was not required for this study because all data used were publicly available, de-identified, and did not involve any direct human or animal experimentation.

Data sharing statement

The datasets used to support the findings of this study are publicly available from The Cancer Genome Atlas (https://portal.gdc.cancer.gov/) and the Clinical Proteomic Tumor Analysis Consortium (https://cptac-data-portal.georgetown.edu/). No additional data are available.

References

[1] He C, Huang Q, Zhong S, Chen LS, Xiao H, Li L. Screening and identifying of biomarkers in early colorectal cancer and adenoma based on genome-wide methylation profiles. World J Surg Oncol 2023;21(1):312. doi:10.1186/s12957-023-03189-1, PMID:37779184.

[2] Wang JM, Hong R, Demicco EG, Tan J, Lazcano R, Moreira AL, *et al*. Deep learning integrates histopathology and proteogenomics at a pan-cancer level. Cell Rep Med 2023;4(9):101173. doi:10.1016/j.xcrm.2023.101173, PMID:37582371.

[3] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, *et al*. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88. doi:10.1016/j.media.2017.07.005, PMID:28778026.

[4] Li D, Wang Z, Liu Y, Zhou M, Xia B, Zhang L, *et al*. Assessing the risk of high-grade squamous intraepithelial lesions (HSIL+) in women with LSIL biopsies: a machine learning-based study. Infect Agent Cancer 2024;19(1):61. doi:10.1186/s13027-024-00625-z, PMID:39639322.

[5] Hu J, Lv H, Zhao S, Lin CJ, Su GH, Shao ZM. Prediction of clinicopathological features, multi-omics events and prognosis based on digital pathology and deep learning in HR(+)/HER2(-) breast cancer. J Thorac Dis 2023;15(5):2528–2543. doi:10.21037/jtd-23-445, PMID:37324098.

[6] Ali H. Artificial intelligence in multi-omics data integration: Advancing precision medicine, biomarker discovery and genomic-driven disease interventions. Int J Sci Res Arch 2023;8(1):1012–1030. doi:10.30574/ijsra.2023.8.1.0189.

[7] Sartori F, Codicè F, Caranzano I, Rollo C, Birolo G, Fariselli P, *et al*. A Comprehensive Review of Deep Learning Applications with Multi-Omics Data in Cancer Research. Genes (Basel) 2025;16(6):648. doi:10.3390/genes16060648, PMID:40565540.

[8] Jing F, Zhu L, Zhang J, Zhou X, Bai J, Li X, *et al*. Multi-omics reveals lactylation-driven regulatory mechanisms promoting tumor progression in oral squamous cell carcinoma. Genome Biol 2024;25(1):272. doi:10.1186/s13059-024-03383-8, PMID:39407253.

[9] Luo Q, Wang J, Chen T, Li J. Pan-cancer multi-omics analysis reveals IQCE as a malignant cell-restricted oncogenic biomarker driving immunosuppression and chemoresistance in cutaneous melanoma. Discov Oncol 2025;16(1):2076. doi:10.1007/s12672-025-03841-0, PMID:41212279.

[10] Wekesa JS, Kimwele M. A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. Front Genet 2023;14:1199087. doi:10.3389/fgene.2023.1199087, PMID:37547471.

[11] Chen L, Li Y, Zhang Z, Yang T, Zeng H. Multi-platform integration of histopathological images and omics data predicts molecular features and prognosis of hepatocellular carcinoma. Front Oncol 2025;15:1591165. doi:10.3389/fonc.2025.1591165, PMID:40766332.

[12] Zhang B, Wan Z, Luo Y, Zhao X, Samayoa J, Zhao W, *et al*. Multimodal integration strategies for clinical application in oncology. Front Pharmacol 2025;16:1609079. doi:10.3389/fphar.2025.1609079, PMID:40910005.

[13] Ren H, Shen C, Hu L, Tang J, Liao Z, Tian W. A Survey of Trends in Biomolecule Recognition for Sensing and Machine Learning Combined with Heterogeneous Information. Curr Bioinform 2025;20:1–18. doi:10.2174/0115748936359132250421053523.

[14] Bao J, Chang C, Zhang Q, Saykin AJ, Shen L, Long Q, *et al*. Integrative analysis of multi-omics and imaging data with incorporation of biological information via structural Bayesian factor analysis. Brief Bioinform 2023;24(2):bbad073. doi:10.1093/bib/bbad073, PMID:36882008.

[15] Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, *et al*. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019;25(7):1054–1056. doi:10.1038/s41591-019-0462-y, PMID:31160815.

[16] Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. Clin Cancer Res 2018;24(6):1248–1259. doi:10.1158/1078-0432.CCR-17-0853, PMID:28982688.

[17] Zhao T, Ren Y, Lu H, Kong Y. Decision level scheme for fusing multiomics and histology slide images using deep neural network for tumor prognosis prediction. Sci Rep 2025;15(1):25479. doi:10.1038/s41598-025-09869-0, PMID:40664732.

[18] Waqas A, Tripathi A, Ramachandran RP, Stewart PA, Rasool G. Multimodal data integration for oncology in the era of deep neural networks: a review. Front Artif Intell 2024;7:1408843. doi:10.3389/frai.2024.1408843, PMID:39118787.

[19] Spanyi A. AI in Precision Oncology. AI in Precision Oncology 2024;1(6):316–317. doi:10.1089/aipo.2024.0045.

[20] Cen X, Lan Y, Zou J, Chen R, Hu C, Tong Y, *et al*. Pan-cancer analysis shapes the understanding of cancer biology and medicine. Cancer Commun (Lond) 2025;45(7):728–746. doi:10.1002/cac2.70008, PMID:40120098.

[21] Huang Z, Li J, Zhou YL, Shi J. Integrated multiomics machine learning and mediated Mendelian randomization investigate the molecular subtypes and prognosis lung squamous cell carcinoma. Transl Lung Cancer Res 2025;14(3):857–877. doi:10.21037/tlcr-24-891, PMID:40248728.

[22] Landwehr GM, Bogart JW, Magalhaes C, Hammarlund EG, Karim AS, Jewett MC. Accelerated enzyme engineering by machine-learning guided cell-free expression. Nat Commun 2025;16(1):865. doi:10.1038/s41467-024-55399-0, PMID:39833164.

[23] Noller K, Botsis T, Camara PG, Ciotti L, Cooper LAD, Goecks J, *et al*. Informatics at the Frontier of Cancer Research. Cancer Res 2025;85(16):2967–2986. doi:10.1158/0008-5472.CAN-24-2829, PMID:40600473.

[24] Ektefaie Y, Yuan W, Dillon DA, Lin NU, Golden JA, Kohane IS, *et al*. Integrative multiomics-histopathology analysis for breast cancer classification. NPJ Breast Cancer 2021;7(1):147. doi:10.1038/s41523-021-00357-y, PMID:34845230.

[25] Lee JS. Exploring cancer genomic data from the cancer genome atlas project. BMB Rep 2016;49(11):607–611. doi:10.5483/bmbrep.2016.49.11.145, PMID:27530686.

[26] Savage SR, Yi X, Lei JT, Wen B, Zhao H, Liao Y, *et al*. Pan-cancer proteogenomics expands the landscape of therapeutic targets. Cell

2024;187(16):4389–4407.e15. doi:10.1016/j.cell.2024.05.039, PMID: 38917788.
[27] Lei JT, Savage SR, Yi X, Wen B, Zhao H, Somes LK, *et al*. Pan-cancer proteogenomics expands the landscape of therapeutic targets. Cancer Res 2023;83(7_Suppl):5726. doi:10.1158/1538-7445.AM2023-5726.
[28] Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, *et al*. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Boston, MA, USA: IEEE; 2009:1107–1110. doi:10.1109/ISBI.2009.5193250.
[29] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15(12):550. doi:10.1186/s13059-014-0550-8, PMID:25516281.
[30] Aburass S, Dorgham O, Al Shaqsi J, Abu Rumman M, Al-Kadi O. Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases. J Imaging Inform Med 2025;38(6):3928–3971. doi:10.1007/s10278-025-01481-y, PMID:40164818.
[31] Hao Y, Cheng C, Li J, Li H, Di X, Zeng X, *et al*. Multimodal Integration in Health Care: Development With Applications in Disease Management. J Med Internet Res 2025;27:e76557. doi:10.2196/76557, PMID:40840463.
[32] Lee SI, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, *et al*. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun 2018;9(1):42. doi:10.1038/s41467-017-02465-5, PMID:29298978.
[33] Ozaki Y, Broughton P, Abdollahi H, Valafar H, Blenda AV. Integrating Omics Data and AI for Cancer Diagnosis and Prognosis. Cancers (Basel) 2024;16(13):2448. doi:10.3390/cancers16132448, PMID:39001510.
[34] Schneider L, Laiouar-Pedari S, Kuntz S, Krieghoff-Henning E, Hekler A, Kather JN, *et al*. Integration of deep learning-based image analysis and genomic data in cancer pathology: A systematic review. Eur J Cancer 2022;160:80–91. doi:10.1016/j.ejca.2021.10.007, PMID:34810047.
35 Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, *et al*. Towards a general-purpose foundation model for computational pathology. Nat Med 2024;30(3):850–862. doi:10.1038/s41591-024-02857-3, PMID:38504018.