Cancer Biol Med 2026. doi: 10.20892/j.issn.2095-3941.2025.0771

ORIGINAL ARTICLE

Multimodal artificial intelligence predicts *PIK3CA* mutation in breast cancer from digital pathology and clinical data: a multicenter study

Jiaxian Miao[1]*, Qi Liu[2]*, Jianing Zhao[1]*, Shishun Fan[1], Shenwen Wang[2], Feng Ye[3], Si Wu[1], Jinze Li[1], Huirui Zhang[1], Meng Zhang[4], Hong Bu[3], Xiao Han[5], Lianghong Teng[4], Yueping Liu[1]

[1]Department of Pathology, Hebei Medical University Fourth Hospital, Shijiazhuang 050011, China; [2]School of Information Engineering, Hebei GEO University, Shijiazhuang 052161, China; [3]Laboratory of Pathology, West China Hospital, Sichuan University, Chengdu 610041, China; [4]Department of Pathology, Xuanwu Hospital, Capital Medical University, Beijing 100053, China; [5]College of Biomedical Engineering, Sichuan University, Chengdu 610065, China

ABSTRACT

Objective: Accurate detection of *PIK3CA* mutations is essential for guiding PI3K-targeted therapies in breast cancer, yet sequencing is not universally accessible, and single-modality prediction models have limited performance. This study developed a multimodal deep learning framework integrating whole-slide imaging (WSI) and structured clinical data to improve mutation prediction.

Methods: A total of 1,047 patients from TCGA and 166 patients from 3 external centers were included. The histopathology model used a transformer-based pretrained encoder (H-optimus-0) and a clustering-constrained attention multiple instance learning (CLAM-SB MIL) classifier to generate WSI-level representations. The clinical model incorporated engineered clinical variables and an extreme gradient boosting (XGBoost) model. A decision-level late fusion strategy (Multimodal *PIK3CA* Model, MPM) combined probabilistic outputs from both branches. Performance was evaluated with the area under the curve (AUC) and secondary metrics. Interpretability was assessed *via* attention heatmaps and shapley additive explanations (SHAP) analysis.

Results: MPM outperformed single-modality models. It achieved an AUC of 0.745 on TCGA and maintained stable performance across external cohorts (0.695, 0.690, and 0.680). SHAP analysis identified molecular subtype as the most influential clinical feature, whereas attention maps highlighted mutation-associated morphological regions.

Conclusions: The developed multimodal framework effectively integrates complementary morphological and clinical information, and provides a robust and generalizable method for predicting *PIK3CA* mutation status. Strong multicenter adaptability and biological interpretability support its potential use as a clinical decision-support tool and an accessible alternative to molecular testing.

KEYWORDS Breast cancer; *PIK3CA* mutation; multimodal artificial intelligence; whole-slide imaging; computational pathology

Introduction

Breast cancer is among the most common malignancies in women worldwide and remains a major global health burden. According to global cancer statistics, breast cancer accounts for nearly 30% of all cancer diagnoses in women and substantially contributes to cancer-related mortality each year[1,2]. Despite advances in diagnosis and treatment, the heterogeneous molecular landscape and complex biological behavior of breast cancer continue to challenge precision oncology[3]. Among the genomic alterations implicated in breast tumorigenesis, mutations within the PI3K/AKT/mTOR (PAM) signaling pathway are among the most prevalent[4,5]. Recent large-scale genomic studies have shown that approximately 62.6% of Chinese patients with breast cancer bear at least one PAM pathway mutation, and *PIK3CA* is the most frequently altered gene. *PIK3CA* mutations occur in as many as 40% of HR+/HER2− tumors and approximately 30% of HER2+ tumors[6].

With the development of PI3K-targeted therapies, *PIK3CA* mutations have emerged as a key therapeutic biomarker. Several novel PI3K inhibitors have demonstrated promising anti-tumor activity[7], and the first-in-class PI3Kα-selective

*These authors contributed equally to this work.
Correspondence to: Yueping Liu, Xiao Han and Lianghong Teng
E-mail: annama@163.com, xiao_han@scu.edu.cn and tenglianghong@xwhosp.org
ORCID ID: https://orcid.org/0000-0002-4582-114X, https://orcid.org/0000-0002-5151-6547 and https://orcid.org/0000-0001-6386-4162
Received November 30, 2025; accepted February 23, 2026.
Available at www.cancerbiomed.org
©2026 The Authors. Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

inhibitor (alpelisib) has been approved by the U.S. Food and Drug Administration and European Medicines Agency for the treatment of patients with *PIK3CA*-mutated HR+/HER2− breast cancer[8]. Consequently, accurate determination of *PIK3CA* mutation status has become increasingly important for treatment decision-making and individualized therapy.

Traditional approaches for detecting *PIK3CA* mutations rely on molecular assays such as PCR and NGS. Although highly accurate, these techniques require advanced laboratory infrastructure, incur substantial cost, and substantially depend on tissue quality; therefore, their application in routine clinical practice has been limited. Deep learning models can now predict key gene mutations, such as *TP53*, *EGFR*, and *KRAS*, directly from routine H&E-stained whole-slide images (WSIs); this major advancement in computational pathology[9] offers a scalable and cost-effective alternative. However, most existing studies have relied solely on single-modality inputs, primarily pathology images, without leveraging complementary clinical information. Multimodal learning, which integrates WSIs with clinical data, has the potential to more comprehensively capture tumor biology and improve mutation prediction accuracy.

Moreover, current Artificial Intelligence (AI) models for predicting *PIK3CA* mutation status were trained predominantly on single-center datasets with limited sample sizes or restricted imaging diversity. Differences in staining protocols, scanning devices, and tissue preparation across institutions often introduce domain shift and substantially limit model robustness in real-world clinical settings. Although recent studies have suggested that incorporating structured clinical variables might enhance prediction performance, large-scale, multi-center, multimodal frameworks specifically designed to predict *PIK3CA* mutations remain lacking.

This study integrated The Cancer Genome Atlas (TCGA) breast cancer dataset with 3 independent multi-center cohorts to develop a decision-level late-fusion multimodal model for predicting *PIK3CA* mutation status (Multimodal *PIK3CA* Model, MPM). By jointly leveraging deep representations from whole-slide histopathology images and structured clinical variables, the model demonstrated robust predictive performance across heterogeneous datasets. MPM effectively capitalizes on the complementary strengths of both modalities in generating stable and generalizable predictions, thereby providing a solid methodological and empirical foundation for subsequent algorithmic refinement, clinical translation, and real-world deployment.

Materials and methods

Data sources and patient selection

The training dataset for this study was obtained from TCGA, which provides multimodal breast cancer data including genomic profiles, clinical information, and WSIs. External validation datasets were collected from 3 independent medical centers: Hebei Medical University Fourth Hospital (HBMU), Beijing Xuanwu Hospital of Capital Medical University (CMU), and West China Hospital of Sichuan University (SCU). The external cohorts consisted of core needle biopsy specimens from patients with breast cancer.

The inclusion criteria were (1) histologically confirmed primary breast cancer; (2) available *PIK3CA* mutation status; (3) high-quality WSIs suitable for downstream analysis; and (4) complete clinical information. The exclusion criteria were (1) absence of *PIK3CA* mutation results; (2) poor-quality WSIs (e.g., severe blurring, artifacts, or incomplete tissue); (3) diagnosis of carcinoma *in situ*; and (4) intraoperative frozen-section slides.

A total of 1,047 patients with 1,112 WSIs were included in the TCGA training cohort. The external validation cohorts consisted of 70 patients with 175 WSIs from HBMU, 52 patients with 52 WSIs from CMU, and 44 patients with 44 WSIs from SCU. Clinical variables included demographic characteristics or pathological biomarkers, such as age, molecular subtype, lymph node status, and American Joint Committee on Cancer (AJCC) stage[10]. All enrolled patients in the external validation cohorts had complete information for the core variables required by the model. External datasets were entirely withheld from model development and used exclusively for independent performance evaluation.

Histopathology image preprocessing and feature extraction

To ensure that only high-quality tissue regions were analyzed, we subjected WSIs to a standardized preprocessing workflow. Tissue segmentation was performed with a combination of the Clustering-constrained Attention Multiple Instance Learning (CLAM)[11] framework and an HSV color-space thresholding approach to exclude background, debris, and low-quality regions. Tile extraction was performed with a fixed tile size of 224 × 224 pixels, and all tiles were stored in HDF5 format.

Five state-of-the-art pathology foundation models were systematically evaluated as feature encoders, representing diverse architectures and pretraining paradigms:

- UNI-V2[12]: Vision Transformer (ViT) models trained *via* DINOv2 self-supervised learning (1,536-d features)
- CTransPath[13]: Hybrid CNN–Swin Transformer encoder trained with contrastive learning (768-d features)
- Virchow2[14]: ViT model combining convolution and global attention to capture multiscale context (2,560-d features)
- H-optimus-0[15]: Efficient pathology-specific ViT encoder (1,536-d features)
- CONCH-V1.5[16]: Vision-language foundation model specialized for pathology (768-d features)

These encoders were benchmarked on the TCGA dataset to determine the optimal feature extractor for downstream WSI-level mutation prediction.

Architecture of the multimodal *PIK3CA* model

Herein, we introduce the MPM, an artificial intelligence framework designed to predict *PIK3CA* mutation status in breast cancer by leveraging heterogeneous multimodal inputs. The MPM uses a decision-level fusion mechanism to combine deep morphological representations from WSIs with clinically relevant variables. The framework is composed of 2 distinct predictive components that jointly form its core:

1. A histopathology model dedicated to processing WSI: We used a multiple instance learning pipeline built upon a pretrained Transformer-based architecture to extract high-resolution morphological patterns associated with *PIK3CA* mutations. The model outputs an independent probability prediction for each case.
2. A clinical model designed to analyze structured clinical variables, including age, molecular subtype, lymph node status, and AJCC stage: We used machine learning algorithms to model the nonlinear relationships between these clinical attributes and *PIK3CA* mutation status, thereby yielding an independent probability prediction for each case.

During the training phase, both models were optimized independently to maximize their respective predictive capabilities. In the inference phase, their output probabilities were combined through our decision-level fusion module, thus yielding a unified and robust prediction of *PIK3CA* mutation status.

Construction of the histopathology model

The histopathology model used a two-stage weakly supervised learning pipeline. The framework first used the pathology foundation model H-optimus-0, pretrained on large-scale histopathology image datasets, to learn deep feature representations. These features were then passed to an attention-based multiple instance learning classifier, CLAM-SB[11], which performed slide-level prediction of *PIK3CA* mutation status.

Given a high-resolution WSI, we first divided it into a sequence of N non-overlapping image patches $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N\}$. Each patch $\mathbf{p}_i$ was then processed through the H-optimus-0 model *via* forward propagation to obtain its high-dimensional deep feature representation.

$$\mathbf{h}_i = f_{\text{H-optimus-0}}(\mathbf{p}_i; \theta_{\text{pre}}), \quad \mathbf{h}_i \in R^d$$

Here, $f_{\text{H-optimus-0}}$ denotes the forward computation function of the pretrained model, θ_{pre} represents its fixed parameters, and d is the dimensionality of the feature vector. This step converted the large-scale WSI into a set of deep feature vectors $H = \{h_1, h_2, \ldots, h_N\}$, and effectively encoded the morphological information of local regions.

Subsequently, the feature set H was fed into the CLAM-SB classification model. This model uses an attention network to compute the importance weight of each image patch with respect to the final classification decision. Its core mechanism involves spatially encoding the features and calculating the corresponding attention scores:

$$e_i = v^T \tanh(Wh_i + b), \quad a_i = \frac{\exp(e_i)}{\sum_{j=1}^{N} \exp(e_j)}$$

Here, $W \in \mathrm{R}^{d \times d}$, $b \in \mathrm{R}^d$, and $v \in \mathrm{R}^d$ are learnable parameters, and a_i denotes the normalized attention score for the i-th image patch. The weighted average of all image patches forms the bag-level representation of the entire WSI:

$$z = \sum_{i=1}^{N} a_i h_i$$

Finally, the WSI-level representation z is passed into a fully connected classification layer to produce the predicted probability of *PIK3CA* mutation:

$$\hat{y}_{\text{path}} = \sigma\left(W_c z + b_c\right)$$

where σ is the sigmoid activation function, and W_c and b_c are the parameters of the classification layer.

This two-stage architecture has an advantage of operating without requiring pixel-level or region-level annotations. Instead, it relies solely on weak WSI-level labels to automatically identify morphologically informative regions associated with genetic mutations and to perform end-to-end prediction. This design provides an efficient and interpretable solution for molecular biomarker detection in digital pathology.

Construction of the clinical model

The clinical model included 4 fundamental variables: age, molecular subtype, lymph node status, and AJCC stage. To construct the clinical data prediction model, we used the extreme gradient boosting (XGBoost) algorithm, which has demonstrated outstanding performance in processing tabular data. By sequentially constructing a series of decision trees to iteratively correct the prediction residuals of preceding models and incorporating regularization terms to effectively control model complexity, XGBoost achieves remarkable predictive accuracy, high computational efficiency, and robust resistance to overfitting.

Given a clinical dataset with n samples, $D = \{(x_i, y_i)\}$, where $x_i \in \mathrm{R}_m$ denotes an m-dimensional feature vector, and $y_i \in \{0, 1\}$ represents the *PIK3CA* mutation status label, the XGBoost model produces its prediction as an additive combination of K decision trees:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

Here, $\hat{y}_i$ denotes the predicted output, f_k represents the k-th decision tree, and F denotes the space of all possible decision trees. The objective function of the model consists of a loss term and a regularization term:

$$L(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \left[\gamma T + (1/2) \lambda \|w\|^2 + \alpha \|w\|_1 \right]$$

Here, $l(y_i, \hat{y}_i)$ denotes the cross-entropy loss function. The regularization term includes T, the number of leaf nodes in the tree; w, the weight of leaf nodes; γ, controlling the tree structure complexity; λ, the L2 regularization coefficient; and α, the L1 regularization coefficient.

Construction of the multimodal *PIK3CA* model

To integrate the complementary information provided by WSI and clinical data for predicting *PIK3CA* mutations in breast cancer, we developed a late decision-level fusion multimodal learning framework, referred to as the MPM (Figure 1). The core idea of this framework is using the prediction outputs from 2 independent branches, the histopathology model and the clinical model, and fusing them through a shallow classifier. Each branch produces a binary probability vector, wherein the 2 elements represent the predicted probabilities of the sample being *PIK3CA* wild-type or mutant, respectively, with a sum equal to 1. The implementation of MPM consists of 2 major stages: training and inference.

1. Training phase

This phase constructs a multimodal fusion training set and trains the logistic regression fusion module. The detailed steps are as follows:

Histopathology model: For the i-th patient in the training set, if the patient has S_i WSIs, each slide is processed independently through the histopathology model to obtain a binary probability vector: $\mathbf{p}_{\text{img}}^{(i,s)} = \left[\mathbf{p}_{\text{img},0}^{(i,s)}, \mathbf{p}_{\text{img},1}^{(i,s)} \right], \quad s = 1, 2, \ldots, S_i.$

The final histopathology model representation for the patient is computed as the arithmetic mean of all slide-level probability vectors:

$$\mathbf{p}_{\text{img}}^{(i)} = \frac{1}{S_i} \sum_{s=1}^{S_i} \mathbf{p}_{\text{img}}^{(i,s)}$$

This aggregation assumes that each slide contributes equally to the final prediction and is designed to reduce random errors from individual slides through ensemble averaging, thereby producing a more robust case-level representation.

Clinical model: For the same patient i, the clinical model directly outputs a binary probability vector based on the structured clinical features:

$$\mathbf{p}_{\text{clin}}^{(i)} = \left[\mathbf{p}_{\text{clin}}^{(i,0)}, \mathbf{p}_{\text{clin}}^{(i,1)} \right].$$

MPM: The multimodal fusion feature for patient i is formed by concatenating $\mathbf{p}_{\text{img}}^{(i)}$ and $\mathbf{p}_{\text{clin}}^{(i)}$:

$$\mathbf{p}_{\text{fusion}}^{(i)} = \mathbf{p}_{\text{img}}^{(i)} \oplus \mathbf{p}_{\text{clin}}^{(i)}.$$

All fusion feature vectors $\left\{ \mathbf{p}_{\text{fusion}}^{(i)} \right\}$ and their corresponding ground-truth labels $\{y^{(i)}\}$ are used to train a logistic regression model, which learns the optimal fusion weights and bias.

2. Inference phase

For a new sample, we first obtain its image-based and clinical-based probability vectors with the trained image and clinical branch models. After the same procedure as in the training phase, the multimodal fusion feature is constructed and is then fed into the trained logistic regression fusion module to produce the final prediction:

$$y_{\text{fusion}} = \text{softmax}\left(W \cdot \mathbf{p}_{\text{fusion}} + b \right),$$

where W and b are the weight matrix and bias vector learned during training. The output y_{fusion} represents the final probability distribution predicted by the MPM.

Experimental settings

All experiments in this study were conducted under a unified experimental setup. The hardware platform was equipped with an NVIDIA GeForce RTX 4090 GPU (24 GB VRAM). The software environment was built on Python 3.8 and PyTorch 2.0.0, with parallel computing acceleration enabled *via* CUDA 11.8. All model development and training were performed in the Ubuntu 20.04 operating system.

To ensure rigor and reproducibility, we used a standardized 5-fold cross-validation scheme. In each fold, patient-level stratified random sampling was used to partition the dataset into training (80%), validation (10%), and test (10%) subsets, ensuring balanced class distributions across splits. The validation set was used for monitoring training progress and early stopping, whereas the test set was reserved for final performance evaluation.

For the histopathology model, the CLAM-SB architecture was used to process WSI features. Input features with a dimensionality of 1,536 were refined and aggregated through a gated attention network. Model weights were initialized with the Xavier method, and Dropout ($P = 0.25$) was applied to enhance generalization. Training was performed with the Adam optimizer with an initial learning rate of 1×10^{-3} and a weight decay of 1×10^{-5}. The model was trained in a whole-slide-per-batch manner for as many as 200 epochs, and early stopping was triggered after 30 epochs without improvement. Key CLAM-SB hyperparameters included a bag-level loss weight of 0.7 and 8 instances sampled per bag (balanced between positive and negative examples). L2 regularization was applied during training to prevent overfitting, and random seeds were fixed to ensure reproducibility.

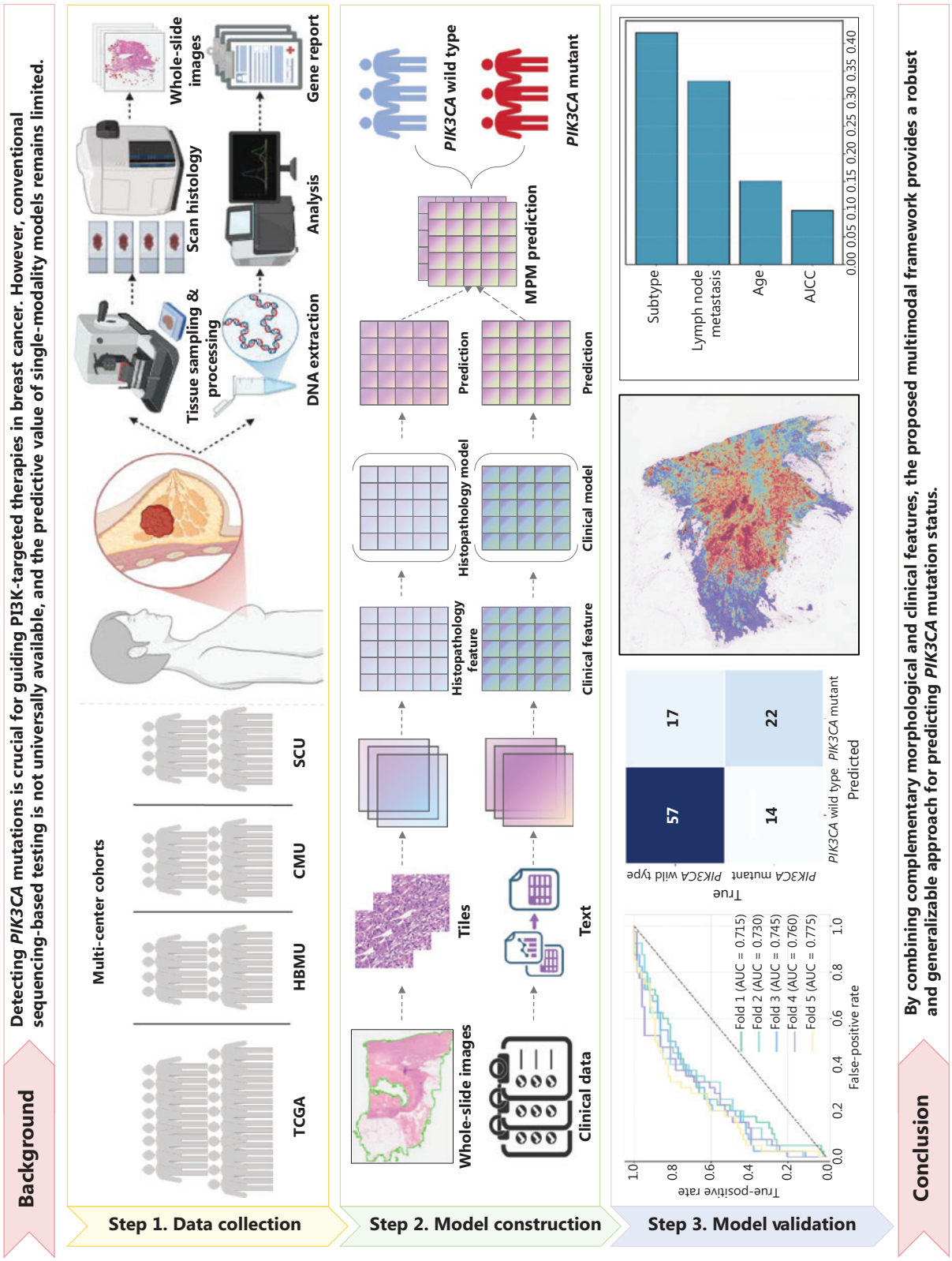


Figure 1 Study flowchart. This study developed a multimodal artificial intelligence framework to predict *PIK3CA* mutations in breast cancer. Step 1: data collection. Multi-center cohorts were assembled from TCGA and 3 external hospitals. Each case underwent standardized tissue processing, histologic scanning, and genomic testing to obtain WSIs, clinical records, and mutation labels. Step 2: model construction. WSIs

were tiled and encoded with a pretrained transformer-based histopathology model, and predictive scores were output *via* the multiple instance learning (MIL) framework. For the clinical model, structured clinical variables were fed into the dedicated clinical model for the derivation of predictive scores. The 2 models generated probabilistic predictions that were fused through the Multimodal *PIK3CA* Model (MPM). Step 3: model validation. Model performance was evaluated with ROC curves, confusion matrices, and attention or SHAP-based interpretability analyses. Overall, the multimodal approach integrated complementary morphological and clinical information to improve the accuracy and generalizability of *PIK3CA* mutation prediction. AJCC, American Joint Committee on Cancer; AUC, area under the curve; CMU, Beijing Xuanwu Hospital of Capital Medical University; HBMU, Hebei Medical University Fourth Hospital; SCU, West China Hospital of Sichuan University; SHAP, Shapley Additive Explanations; TCGA, The Cancer Genome Atlas; WSI, whole-slide image.

For the clinical model, the XGBoost ensemble learning model was used to process features including age, lymph node metastasis status, molecular subtype, and AJCC stage. The model consisted of 100 decision trees with a maximum depth of 3, a learning rate of 0.1, and subsampling ratios of 80% for both instances and features, to mitigate overfitting. L1 ($\alpha = 0.1$) and L2 ($\lambda = 1.0$) regularization were applied. To address class imbalance, the model automatically adjusted sample weights based on the positive-negative ratio.

Throughout the study, all random partitions and initializations were controlled by a fixed random seed to ensure stability and comparability of results. All training procedures were executed in the described hardware and software environment, thereby providing a reliable foundation for subsequent internal validation and external generalization assessments.

Evaluation metrics

The primary evaluation metric in this study was the area under the curve (AUC). In addition, we used the accuracy (ACC), F1 score (F1), sensitivity (SEN), and specificity (SPEC) to provide a more comprehensive assessment of model performance. All metrics were evaluated with the bootstrap method (1,000 resamples) to calculate 95% confidence intervals (CIs), thereby quantifying the statistical stability of the evaluation results. To assess the significance of differences in AUC among models, we used the DeLong test. Furthermore, the Platt Scaling method was applied to assess the consistency between model-predicted probabilities and actual observed probabilities, and calibration curves were plotted accordingly. Decision curve analysis (DCA) was applied to quantify the clinical net benefit of the model across various thresholds. This comprehensive approach ensured balanced consideration of the model's discriminative ability, calibration performance, and clinical utility. To ensure strict data independence, we performed dataset partition at the patient level, such that all WSIs originating from the same patient were assigned exclusively to the training, validation, or test set. This process prevented any information leakage across folds and ensured fair evaluation.

To further assess the generalization capability of the developed method, we conducted external testing with independent datasets collected from multiple medical centers. These external evaluations enabled a more comprehensive and rigorous examination of the model's robustness across heterogeneous clinical environments.

Attention-based visualization of histopathology model decisions

To interpret the decision-making process of the histopathology model and identify the morphological features most critical for *PIK3CA* mutation prediction, we generated heatmap visualizations based on the attention weights produced by the model. These visualizations revealed the microscopic pathological structures focused on by the model when determining mutation status.

Specifically, for a given WSI, the pretrained vision Transformer architecture assigns a normalized attention score to each image patch. This score, derived from the softmax normalization operation within the attention mechanism, quantifies the relative contribution of each patch to the final prediction of *PIK3CA* mutation status. The normalized attention values are then mapped back to their corresponding spatial locations on the original slide and upsampled to reconstruct a full-resolution heatmap. In the resulting heatmap, regions with high attention scores are highlighted in warm colors (e.g., red), representing tissue areas that the model identifies as most predictive of *PIK3CA* mutation, such as specific tumor cell arrangements, degrees of nuclear atypia, or patterns of stromal reaction.

This visualization approach transforms the internal decision mechanisms of the model into visual cues that can be readily interpreted by pathologists, thereby enhancing the transparency of the deep learning framework. More importantly, it enables verification of whether the regions highlighted by the model correspond to established pathological knowledge associated with *PIK3CA* mutations, thus establishing a reliable pathway from computational decision-making to clinical interpretation.

SHAP analysis of clinical feature contributions

To gain deeper insight into the decision-making process of the XGBoost model and identify the clinical features most critical for predicting *PIK3CA* mutations, we used the SHapley Additive exPlanations (SHAP) method for interpretability

analysis. SHAP, which is based on Shapley value theory from cooperative game theory, assigns a unique contribution value to each feature for each individual prediction, thereby providing consistent and reliable explanations at both global and local levels. For a given sample x and model f, the SHAP value ϕ^j represents the contribution of feature j to the prediction $f(x)$, computed over all possible subsets of features:

$$f_j(f,x) = \sum_{S\subseteq\{1,\ldots,m\}\setminus\{j\}} \frac{|S|!(m-|S|-1)!}{m!}[f(S\cup\{j\}) - f(S)]$$

Here, S denotes a subset of features not including feature j, $f(S)$ represents the model prediction with only the feature subset S, and m is the total number of features. By aggregating the absolute SHAP values across all samples, we obtain the global importance ranking for each feature:

$$I_j = \frac{1}{n}\sum_{i=1}^{n}|\phi_j^{(i)}|$$

where I_i denotes the global importance of feature j, and $\phi_j^{(i)}$ is the SHAP value of feature j for sample i.

Ethical approval

All digital pathology slides were de-identified according to the DICOM standard (with removal of patient ID, hospital ID, and date of birth). Structured clinical data were pseudonymized by replacing identifiers with unique codes. Access to raw data was restricted to the core research team, with encrypted storage and role-based access controls. This study was approved by the Ethics Committee of Hebei Medical University Fourth Hospital (approval No. 2025KT033). The requirement for informed consent was waived, because of the retrospective nature of the study.

Results

Patient cohorts and dataset composition

On the basis of predefined inclusion and exclusion criteria, all cases from TCGA and the 3 external clinical centers were systematically screened. The TCGA cohort included 1,047 patients with 1,112 WSIs, each with complete *PIK3CA* mutation status, high-quality WSIs, and full clinical information. HBMU (70 patients and 175 WSIs), CMU (52 patients and 52 WSIs), and SCU (44 patients and 44 WSIs) were the external validation cohorts. All included cases were primary breast cancers. Patients with carcinoma *in situ*, recurrent or metastatic disease, poor-quality WSIs, missing mutation data, or frozen-section slides were excluded. The final datasets and clinicopathologic characteristics are summarized in **Figure 2** and **Table 1**.

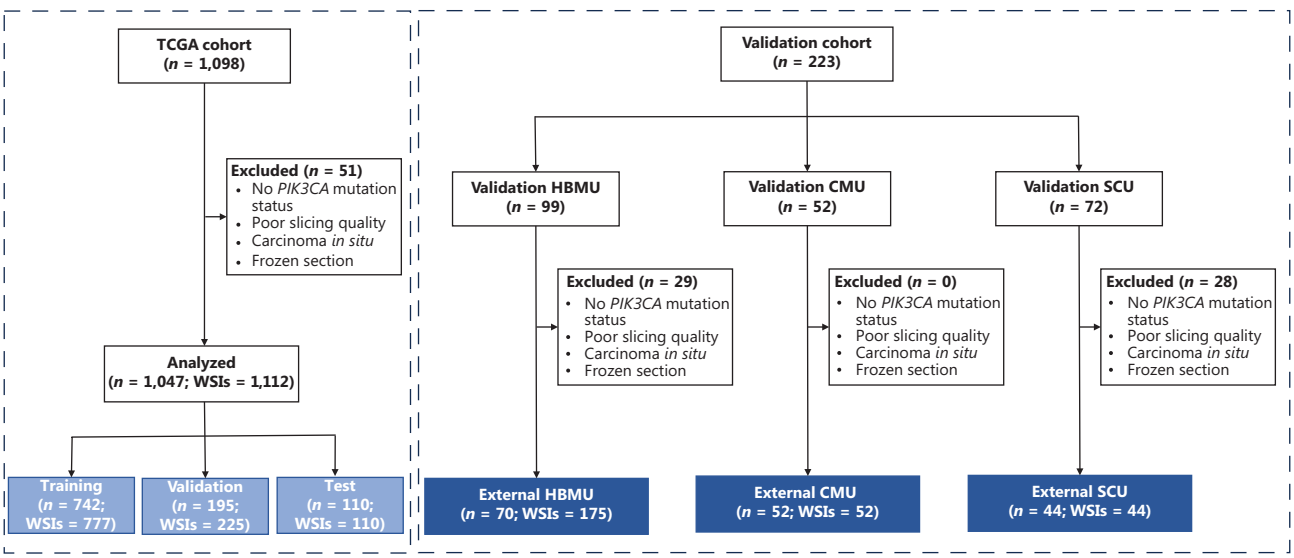


Figure 2 Study cohort construction and data flow diagram. This flowchart summarizes the inclusion and exclusion criteria across the TCGA cohort and 3 external clinical cohorts (HBMU, CMU, and SCU). After application of exclusion criteria including missing *PIK3CA* mutation status, poor image quality, carcinoma *in situ*, or frozen sections, eligible WSIs and clinical data were retained for analysis. The TCGA dataset was further divided into training, validation, and internal test sets. The 3 external cohorts were used for independent external validation. CMU, Beijing Xuanwu Hospital of Capital Medical University; HBMU, Hebei Medical University Fourth Hospital; SCU, West China Hospital of Sichuan University; TCGA, The Cancer Genome Atlas; WSIs, whole-slide images.

Table 1 Clinicopathologic characteristics of the TCGA and external validation cohorts

	TCGA set (*n* = 1,047)	HBMU set (*n* = 70)	CMU set (*n* = 52)	SCU set (*n* = 44)
Age, years	58.420 ± 13.246	50.871 ± 11.178	57.404 ± 12.069	47.045 ± 8.747
Lymph node metastasis, n (%)				
Yes	477 (45.559)	64 (91.428)	10 (19.231)	19 (43.182)
No	451 (43.075)	6 (8.572)	42 (80.769)	25 (56.818)
Unknown	119 (11.366)	0 (0.000)	0 (0.000)	0 (0.000)
Subtype, n (%)				
Basal like	167 (15.950)	19 (27.143)	0 (0.000)	4 (9.091)
HER2 enriched	76 (7.259)	28 (40.000)	0 (0.000)	6 (13.636)
Luminal A	478 (45.654)	1 (1.428)	41 (78.846)	11 (25.000)
Luminal B	195 (18.625)	22 (31.429)	11 (21.154)	23 (52.273)
Unknown	131 (12.512)	0 (0.000)	0 (0.000)	0 (0.000)
AJCC, n (%)				
1.0	196 (18.720)	0 (0.000)	33 (63.462)	0 (0.000)
2.0	598 (57.116)	30 (42.857)	18 (34.615)	11 (25.000)
3.0	235 (22.445)	34 (48.572)	1 (1.923)	33 (75.000)
4.0	18 (1.719)	6 (8.571)	0 (0.000)	0 (0.000)
Label, n (%)				
PIK3CA mutant	319 (30.468)	31 (44.286)	26 (50.000)	21 (47.727)
PIK3CA wild type	728 (69.532)	39 (55.714)	26 (50.000)	23 (52.273)

AJCC, American Joint Committee on Cancer; CMU, Beijing Xuanwu Hospital of Capital Medical University; HBMU, Hebei Medical University Fourth Hospital; SCU, West China Hospital of Sichuan University; SD, standard deviation; TCGA, The Cancer Genome Atlas.

Table 2 Performance comparison of various pretrained feature extractors on the TCGA test set

Pre-encoder	AUC (95% CI)	ACC (95% CI)	F1 (95% CI)	SEN (95% CI)	SPEC (95% CI)	*P*
Virchow2	0.635 (0.570–0.700)	0.628 (0.578–0.678)	0.432 (0.370–0.494)	0.441 (0.355–0.527)	0.728 (0.670–0.786)	<0.001
CONCH-V1.5	0.673 (0.615–0.731)	0.655 (0.607–0.703)	0.478 (0.420–0.536)	0.488 (0.405–0.571)	0.745 (0.690–0.800)	<0.001
CTransPath	0.681 (0.625–0.737)	0.665 (0.619–0.711)	0.495 (0.440–0.550)	0.505 (0.425–0.585)	0.772 (0.720–0.824)	0.015
UNI-V2	0.698 (0.645–0.751)	0.670 (0.628–0.712)	0.515 (0.465–0.565)	0.540 (0.470–0.610)	0.755 (0.705–0.805)	0.045
H-optimus-0	0.715 (0.655–0.775)	0.682 (0.597–0.767)	0.532 (0.447–0.617)	0.561 (0.461–0.661)	0.768 (0.668–0.868)	NA

ACC, accuracy; AUC, area under the curve; CI, confidence interval; F1, F1 score; NA, not available; SEN, sensitivity; SPEC, specificity; TCGA, The Cancer Genome Atlas.

Performance of the histopathology model

Within the TCGA dataset, we conducted a systematic benchmarking analysis of 5 advanced pathology foundation models serving as pretrained encoders in the CLAM-SB architecture (performance metrics summarized in **Table 2**). Among the evaluated encoders, H-optimus-0 achieved the best performance, with an AUC of 0.715 (95% CI: 0.665–0.775) and ACC of 0.682 (95% CI: 0.597–0.767). UNI-V2 ranked second (AUC: 0.698), whereas Virchow2 had the weakest performance (AUC: 0.635). Statistical analysis (DeLong test) revealed that, compared with the top-performing H-optimus-0 model, all other evaluated models exhibited statistically significant performance disadvantages (all $P < 0.05$). Because of its comprehensive superior performance across all evaluation metrics, H-optimus-0 was selected as the feature encoder for the final histopathology branch.

The trained image model was subsequently evaluated on the external multi-center datasets (**Figure 3** and **Table 4**). The

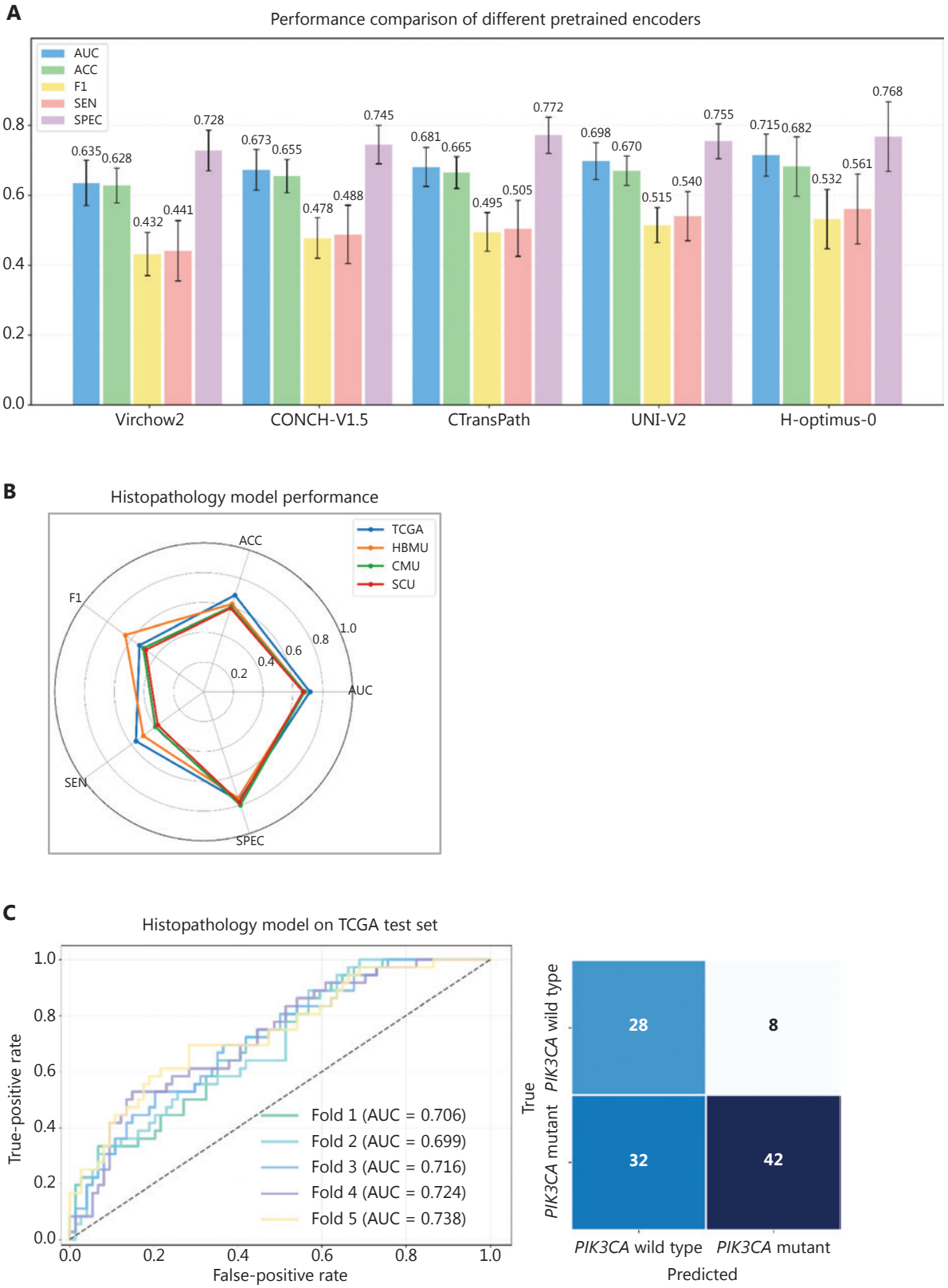


Figure 3 Continued

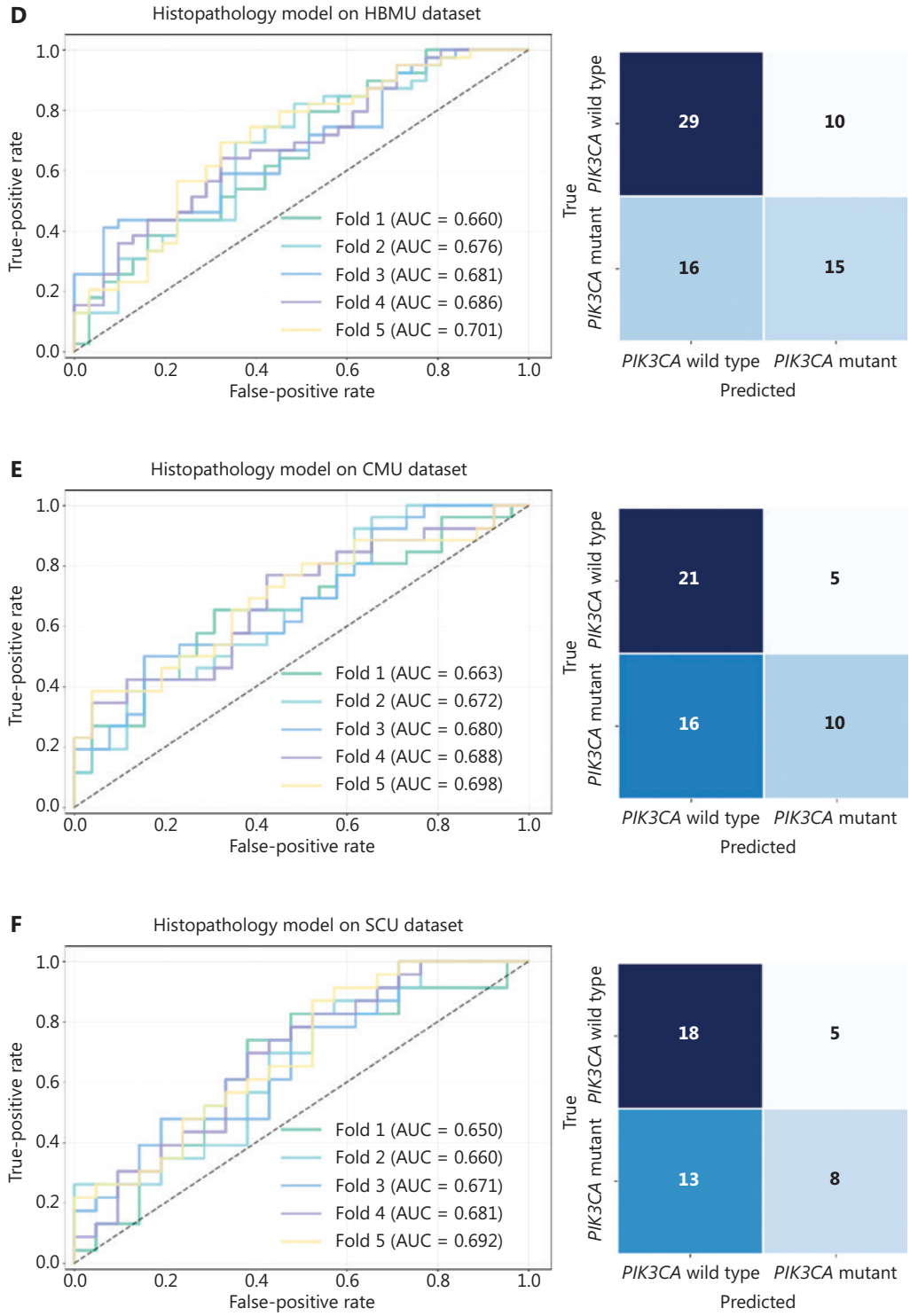


Figure 3 Performance of the histopathology model across datasets. (A) Comparison of prediction performance with various pretrained encoders. (B) Radar plots summarizing AUC, ACC, F1, SEN, and SPEC across centers. (C–F) ROC curves with 5-fold cross-validation results and corresponding confusion matrices of the histopathology model in the (C) TCGA test set, (D) HBMU dataset, (E) CMU dataset, and (F) SCU dataset. ACC, accuracy; AUC, area under the curve; CI, confidence interval; CMU, Beijing Xuanwu Hospital of Capital Medical University; F1, F1 score; HBMU, Hebei Medical University Fourth Hospital; ROC, receiver operating characteristic; SCU, West China Hospital of Sichuan University; SEN, sensitivity; SPEC, specificity; TCGA, The Cancer Genome Atlas.

Table 3 Comparison of machine-learning algorithms for the clinical model

Model	AUC (95% CI)	ACC (95% CI)	F1 (95% CI)	SEN (95% CI)	SPEC (95% CI)	*P*
k-NN	0.534 (0.457–0.612)	0.630 (0.572–0.689)	0.749 (0.704–0.794)	0.796 (0.726–0.866)	0.252 (0.163–0.342)	<0.001
LR	0.558 (0.493–0.624)	0.546 (0.513–0.580)	0.622 (0.592–0.652)	0.537 (0.506–0.568)	0.567 (0.498–0.636)	<0.001
GB	0.681 (0.635–0.728)	0.665 (0.614–0.717)	0.780 (0.747–0.813)	0.855 (0.810–0.900)	0.233 (0.111–0.354)	0.0817
RF	0.681 (0.618–0.743)	0.601 (0.548–0.654)	0.636 (0.567–0.706)	0.506 (0.425–0.587)	0.819 (0.792–0.845)	0.3022
XGBoost	0.694 (0.647–0.741)	0.619 (0.561–0.678)	0.674 (0.610–0.738)	0.570 (0.494–0.646)	0.732 (0.688–0.777)	NA

ACC, accuracy; AUC, area under the receiver operating characteristic curve; CI, confidence interval; F1, F1 score; GB, gradient boosting; k-NN, k-nearest neighbors; LR, logistic regression; NA, not available; RF, random forest; SEN, sensitivity; SPEC, specificity.

results demonstrated stable generalization. The HBMU cohort had an AUC of 0.680 (95% CI: 0.540–0.800) and ACC of 0.620. The CMU cohort had an AUC of 0.680 (95% CI: 0.550–0.810) and ACC of 0.600. The SCU cohort had an AUC of 0.670 (95% CI: 0.520–0.820) and ACC of 0.590. Overall, the histopathology model maintained consistent performance across centers, thus supporting its robustness against staining and scanning variability.

Performance of the clinical model

In the clinical model, 5 machine-learning algorithms were benchmarked: k-nearest neighbors (k-NN), logistic regression (LR), gradient boosting (GB), random forest (RF), and XGBoost. XGBoost outperformed all other models on the TCGA dataset, achieving an AUC of 0.694 (95% CI: 0.647–0.741) and ACC of 0.619 (95% CI: 0.561–0.678) (**Table 3**). Statistical analysis (DeLong test) confirmed the significant performance advantage of XGBoost over k-NN and LR ($P < 0.05$). GB and RF yielded comparable AUC values (both 0.681), whereas k-NN exhibited the lowest predictive performance (AUC = 0.534). Because of its superior and statistically significant predictive capability, XGBoost was selected as the classifier for the clinical model.

To assess the model's true generalization ability, we directly applied the trained clinical model to 3 completely independent external validation cohorts. The model demonstrated stable performance across all external cohorts (**Figure 4** and **Table 4**). In the HBMU cohort, it achieved an AUC of 0.660 (95% CI: 0.560–0.760), specificity of 0.680, and sensitivity of 0.600; in the CMU cohort, the AUC was 0.640 (95% CI: 0.520–0.760), and the model exhibited balanced performance characteristics; and in the SCU cohort, the AUC was 0.650 (95% CI: 0.530–0.770). These results consistently indicated the model's robust discriminative ability across multiple external datasets.

Performance of the multimodal *PIK3CA* model

Using a decision-level late fusion strategy, the MPM integrates the predictive outputs from both the histopathology and clinical models. MPM achieved the best overall performance on the internal TCGA test set, with an AUC of 0.745 (95% CI: 0.715–0.775), ACC of 0.700 (95% CI: 0.660–0.740), F1 of 0.640 (95% CI: 0.600–0.680), SEN of 0.615 (95% CI: 0.570–0.660), and SPEC of 0.775 (95% CI: 0.740–0.810) (**Figure 5** and **Table 4**). Significance testing (DeLong test) confirmed the MPM's statistically superior performance to those of the histopathology model ($P = 0.038$) and clinical model ($P < 0.001$).

On the external validation cohorts, MPM consistently maintained robust and balanced predictive capability. In the HBMU cohort, it attained an AUC of 0.695 (95% CI: 0.615–0.775) and ACC of 0.645 (95% CI: 0.580–0.710). In the CMU cohort, the model achieved an AUC of 0.690 (95% CI: 0.610–0.770) and ACC of 0.615 (95% CI: 0.535–0.695). For the SCU cohort, the MPM yielded an AUC of 0.680 (95% CI: 0.600–0.760) and ACC of 0.620 (95% CI: 0.540–0.700). Across all external centers, the MPM demonstrated stable specificity (0.740–0.760) and provided a more balanced trade-off between sensitivity and specificity than either single-modality model, with no observed statistically significant performance degradation (all $P > 0.05$ *vs.* single modalities in external sets).

Calibration curve analysis demonstrated that the predictive probabilities generated by the MPM were in good agreement with the actual event rates across both the internal TCGA cohort and all external validation cohorts. DCA further validated the clinical utility of the model. Within the threshold probability range of 0.020–0.500, the MPM yielded significantly greater net benefits than the 2 conventional strategies of universal sequencing and no sequencing. These findings underscored the advantages and translational potential of the multimodal fusion strategy in clinical practice.

Overall, these results indicated that, by leveraging complementary morphological features from WSIs and structured clinical data, the MPM substantially improved the predictive performance on the internal cohort and maintained strong, generalizable performance across heterogeneous external centers, thereby supporting the efficacy and robustness of the multimodal fusion strategy.

Table 4 Performance evaluation of single-modal *vs.* multi-modal models

	Histopathology model	Clinical model	MPM
TCGA test set			
AUC (95% CI)	0.715 (0.655–0.775)	0.694 (0.647–0.741)	0.745 (0.715–0.775)
ACC (95% CI)	0.682 (0.597–0.767)	0.619 (0.561–0.678)	0.700 (0.660–0.740)
F1 (95% CI)	0.532 (0.447–0.617)	0.674 (0.610–0.738)	0.640 (0.600–0.680)
SEN (95% CI)	0.561 (0.461–0.661)	0.570 (0.494–0.646)	0.615 (0.570–0.660)
SPEC (95% CI)	0.768 (0.668–0.868)	0.732 (0.688–0.777)	0.775 (0.740–0.810)
P	0.038	<0.001	NA
HBMU			
AUC (95% CI)	0.680 (0.540–0.800)	0.660 (0.560–0.760)	0.695 (0.615–0.775)
ACC (95% CI)	0.620 (0.500–0.730)	0.630 (0.530–0.730)	0.645 (0.580–0.710)
F1 (95% CI)	0.650 (0.540–0.750)	0.650 (0.550–0.750)	0.660 (0.610–0.710)
SEN (95% CI)	0.500 (0.380–0.620)	0.600 (0.500–0.700)	0.575 (0.500–0.650)
SPEC (95% CI)	0.750 (0.600–0.900)	0.680 (0.580–0.780)	0.740 (0.665–0.815)
P	0.550	0.250	NA
CMU			
AUC (95% CI)	0.680 (0.550–0.810)	0.640 (0.520–0.760)	0.690 (0.610–0.770)
ACC (95% CI)	0.600 (0.460–0.740)	0.600 (0.480–0.720)	0.615 (0.535–0.695)
F1 (95% CI)	0.500 (0.360–0.640)	0.620 (0.500–0.740)	0.590 (0.510–0.670)
SEN (95% CI)	0.400 (0.260–0.540)	0.580 (0.460–0.700)	0.525 (0.435–0.615)
SPEC (95% CI)	0.800 (0.660–0.940)	0.650 (0.530–0.770)	0.760 (0.675–0.845)
P	0.550	0.250	NA
SCU			
AUC (95% CI)	0.670 (0.520–0.820)	0.650 (0.530–0.770)	0.680 (0.600–0.760)
ACC (95% CI)	0.590 (0.440–0.740)	0.620 (0.500–0.740)	0.620 (0.540–0.700)
F1 (95% CI)	0.480 (0.330–0.630)	0.640 (0.520–0.760)	0.605 (0.525–0.685)
SEN (95% CI)	0.380 (0.230–0.530)	0.590 (0.470–0.710)	0.535 (0.445–0.625)
SPEC (95% CI)	0.780 (0.630–0.930)	0.670 (0.550–0.790)	0.755 (0.670–0.840)
P	0.480	0.320	NA

ACC, accuracy; AUC, area under the curve; CI, confidence interval; CMU, Beijing Xuanwu Hospital of Capital Medical University; F1, F1 score; HBMU, Hebei Medical University Fourth Hospital; SCU, West China Hospital of Sichuan University; SEN, sensitivity; SPEC, specificity; TCGA, The Cancer Genome Atlas.

Attention heatmap interpretation of the histopathology model

Attention heatmaps derived from the CLAM-SB model highlighted the WSI regions contributing most strongly to *PIK3CA* mutation prediction. Attention heatmaps were generated by the histopathology model across the TCGA, CMU, SCU, and HBMU datasets (**Figure 6**). High-attention regions (in red) localized predominantly to tumor-rich areas characterized by densely packed malignant cells or prominent glandular/tubular structures, whereas low-attention regions (in blue) typically corresponded to normal tissue or non-informative background areas. These patterns indicated the model's accurate focus on diagnostically relevant histomorphologic features,

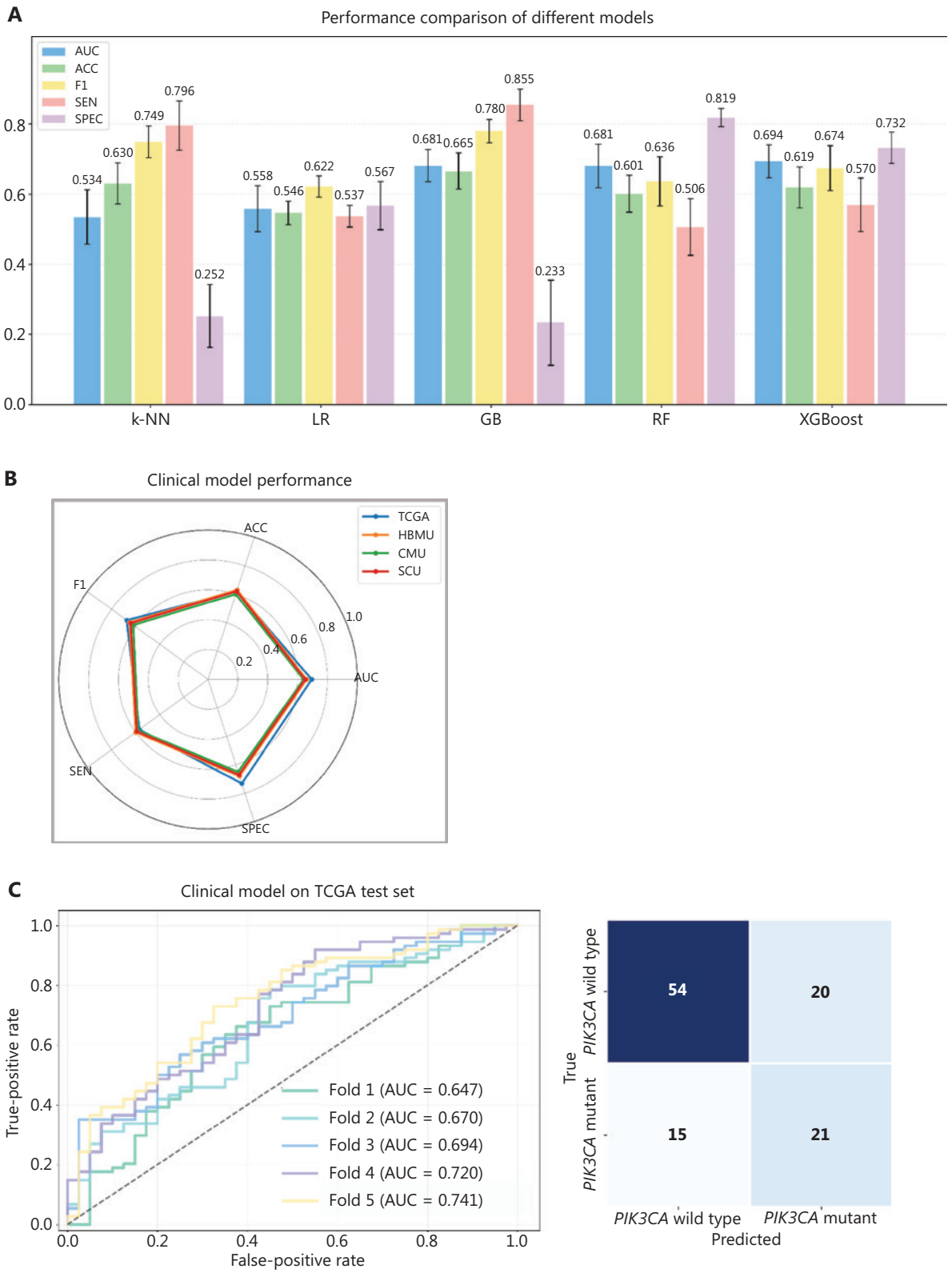


Figure 4 Continued

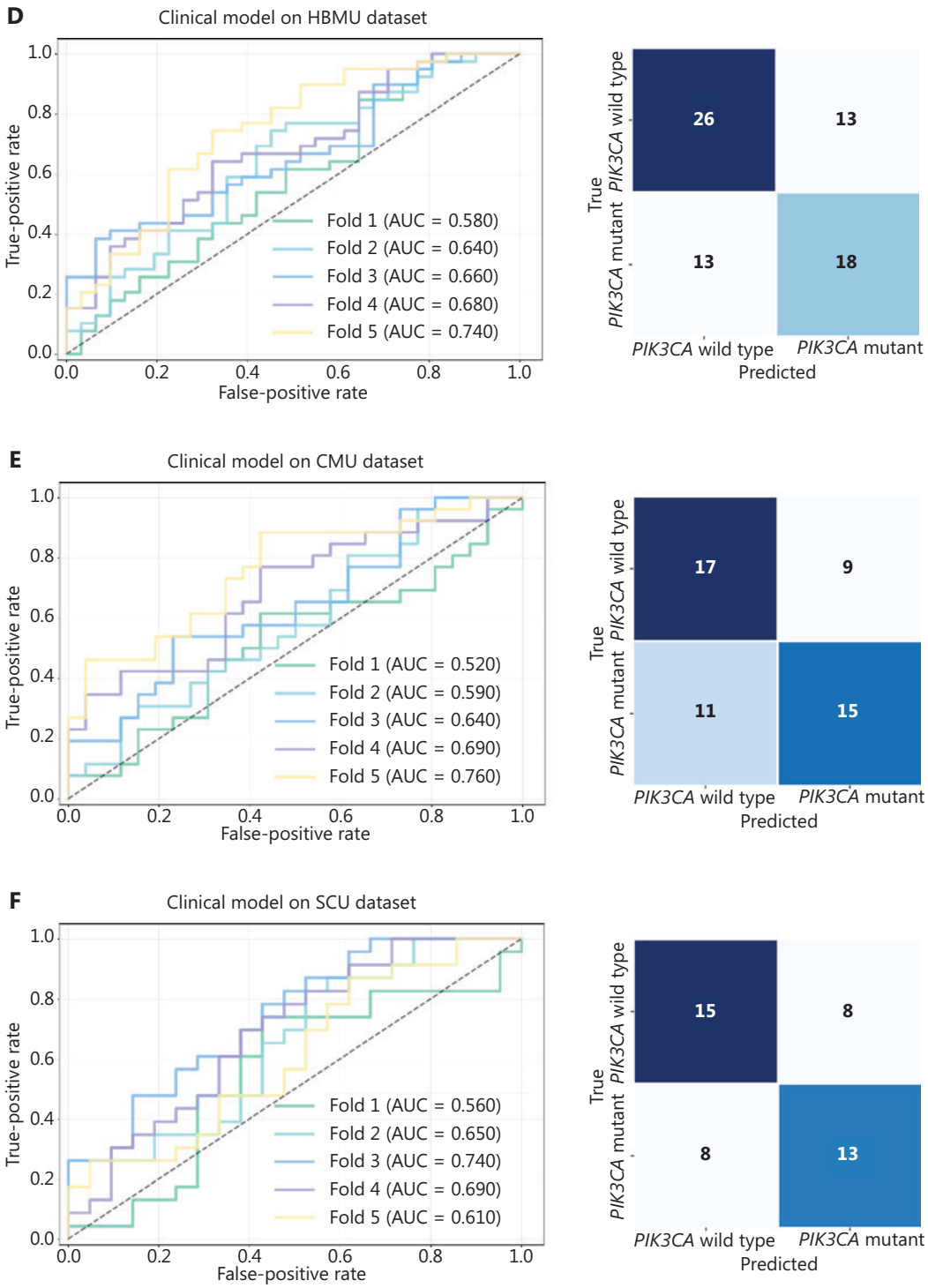


Figure 4 Performance of the clinical model across datasets. (A) Comparison across multiple machine learning classifiers. (B) Radar plots summarizing model performance across datasets. (C–F) ROC curves with 5-fold cross-validation results and corresponding confusion matrices of the clinical model in the (C) TCGA test set, (D) HBMU dataset, (E) CMU dataset, and (F) SCU dataset. ACC, accuracy; AUC, area under the curve; CI, confidence interval; CMU, Beijing Xuanwu Hospital of Capital Medical University; F1, F1 score; HBMU, Hebei Medical University Fourth Hospital; ROC, receiver operating characteristic; SCU, West China Hospital of Sichuan University; SEN, sensitivity; SPEC, specificity; TCGA, The Cancer Genome Atlas.

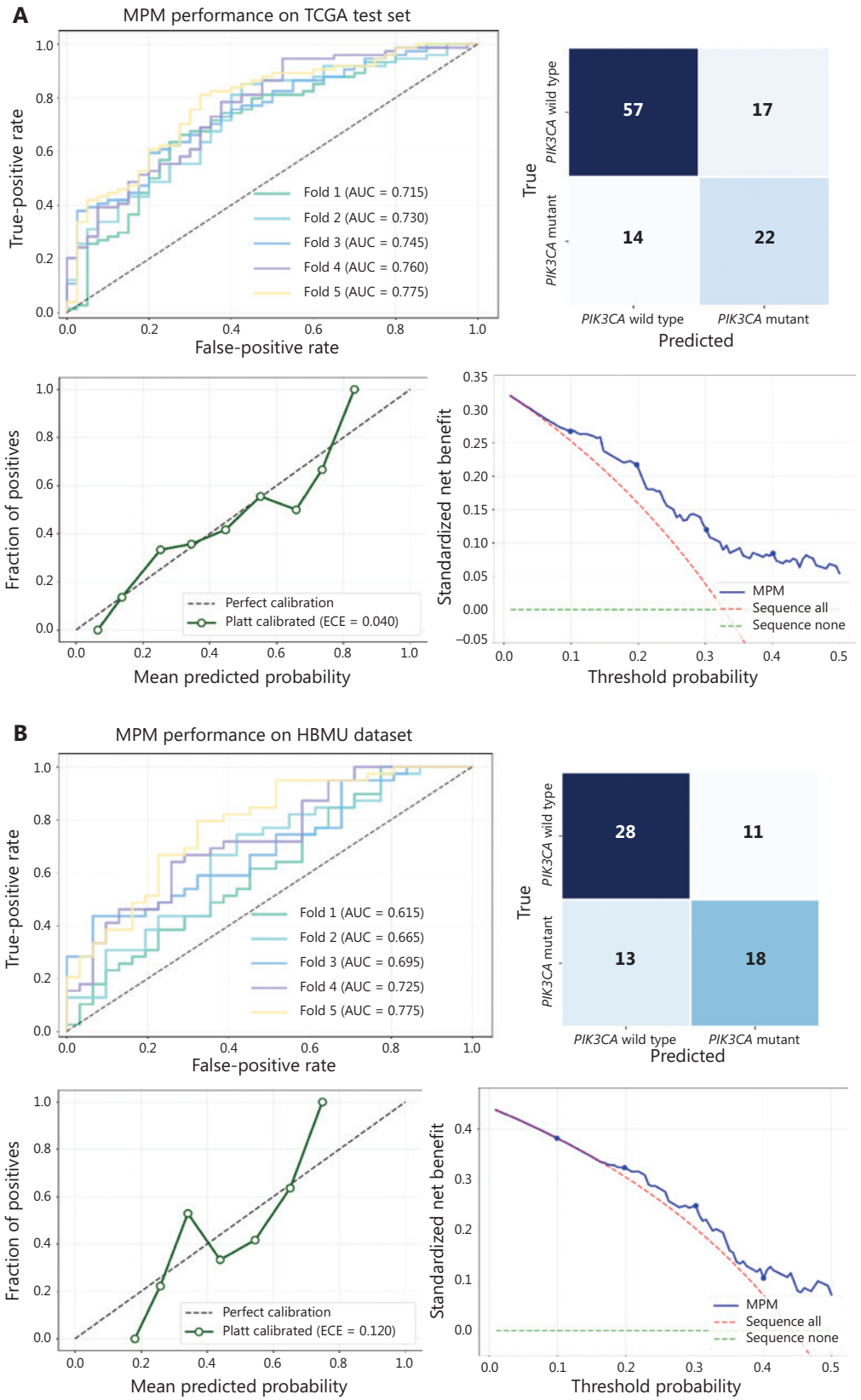


Figure 5 Continued

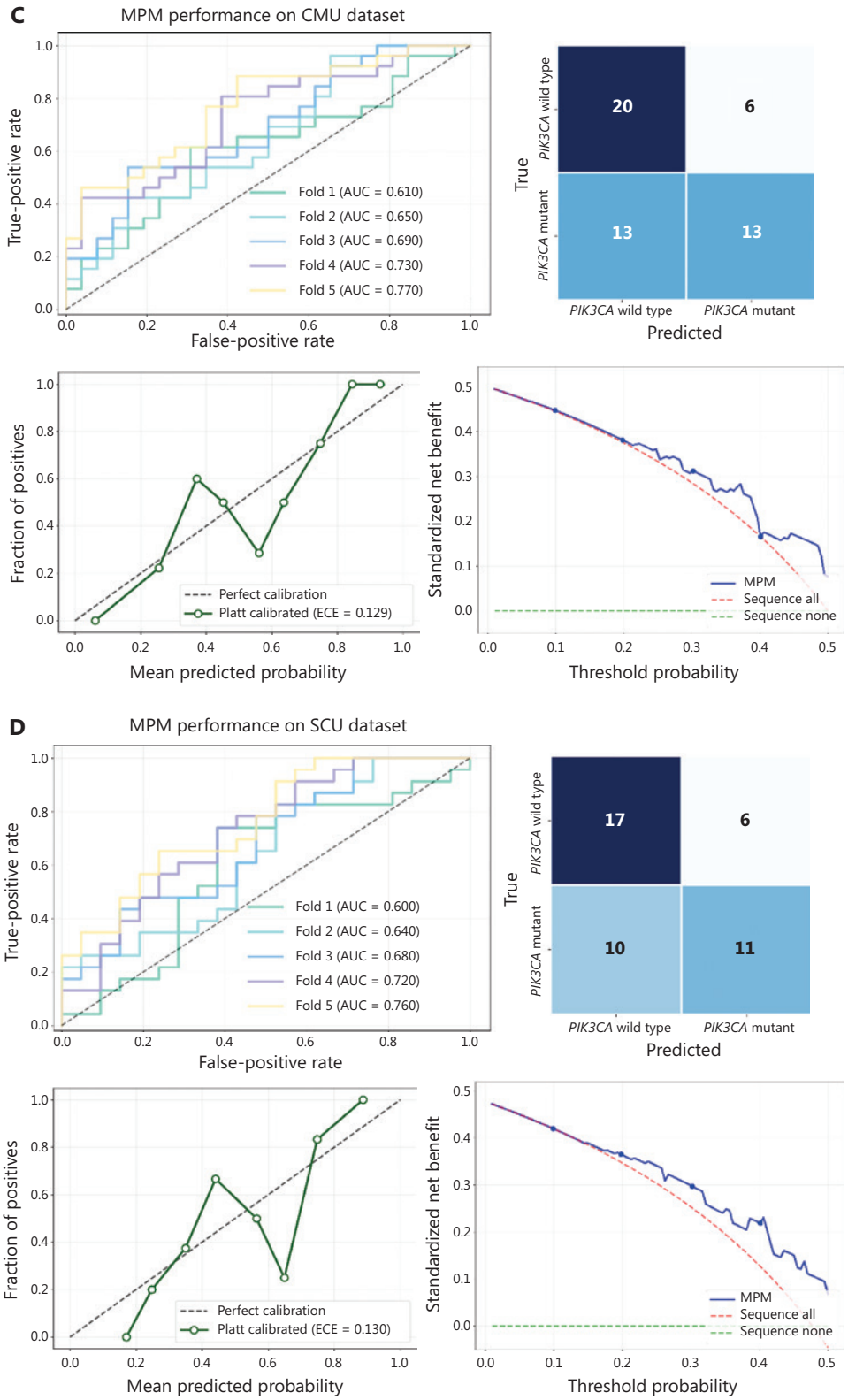


Figure 5 Performance of the multimodal *PIK3CA* Model (MPM). ROC curves, confusion matrices, calibration curves, and DCA results of the MPM in 4 independent cohorts: (A) TCGA test set, (B) HBMU dataset, (C) CMU dataset, and (D) SCU dataset. AUC, area under the curve; CMU, Beijing Xuanwu Hospital of Capital Medical University; DCA, decision curve analysis; HBMU, Hebei Medical University Fourth Hospital; MPM, multimodal *PIK3CA* model; SCU, West China Hospital of Sichuan University; TCGA, The Cancer Genome Atlas.

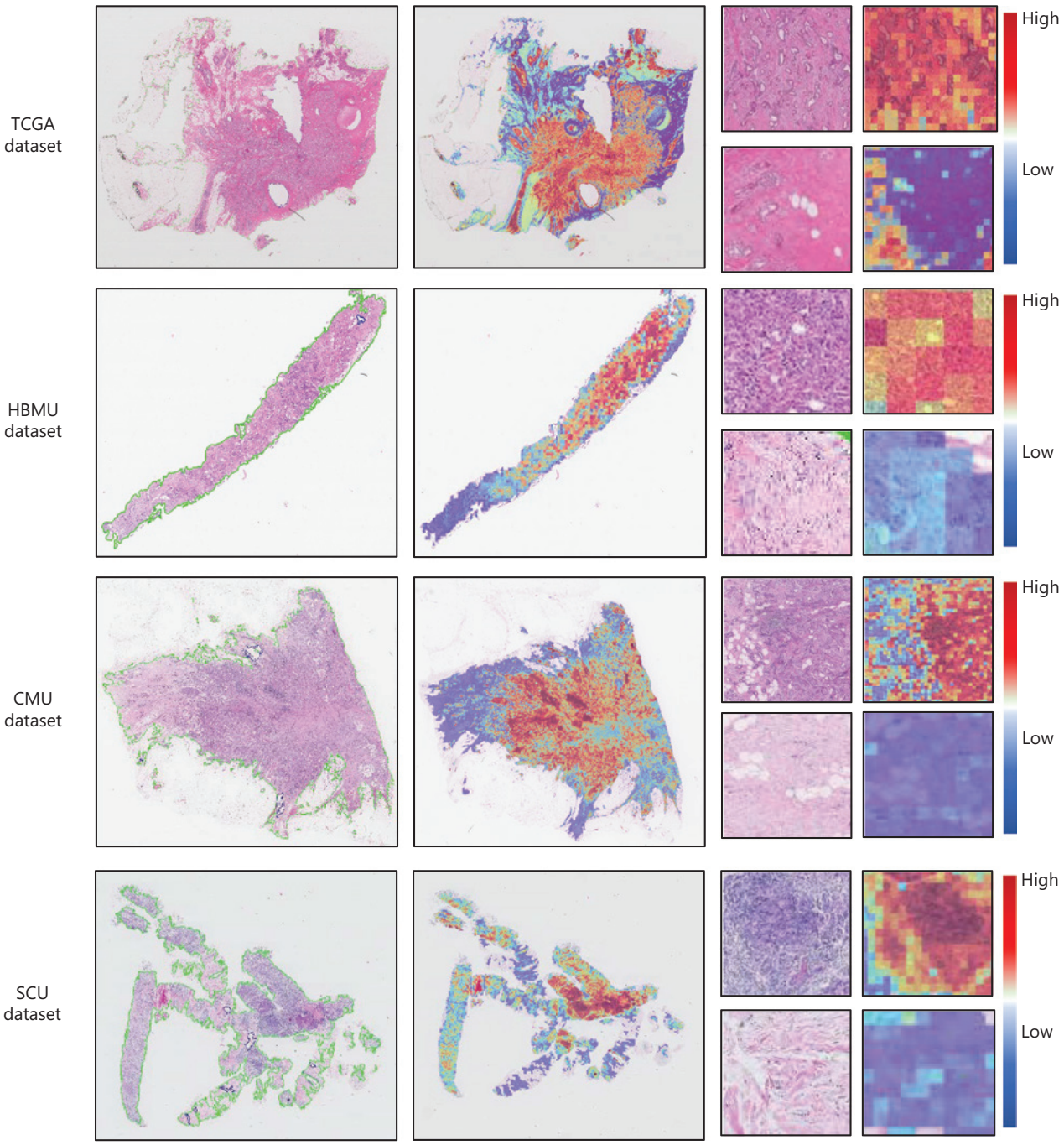


Figure 6 Attention heatmaps of WSI regions contributing to *PIK3CA* mutation prediction. Representative WSIs from TCGA and 3 external cohorts are shown with attention overlays. High-attention regions (red) highlight morphologic areas strongly contributing to mutation prediction. Adjacent panels show magnified patches and corresponding attention heatmaps, reflecting model interpretability and center-to-center robustness. CMU, Beijing Xuanwu Hospital of Capital Medical University; HBMU, Hebei Medical University Fourth Hospital; SCU, West China Hospital of Sichuan University; TCGA, The Cancer Genome Atlas; WSI, whole-slide image.

thereby achieving automatic identification of key regions within WSIs. Furthermore, the consistency of heatmap distributions across datasets from different centers highlighted the robustness and generalizability of the developed image-branch attention mechanism and underscored its strong interpretability and cross-domain adaptability.

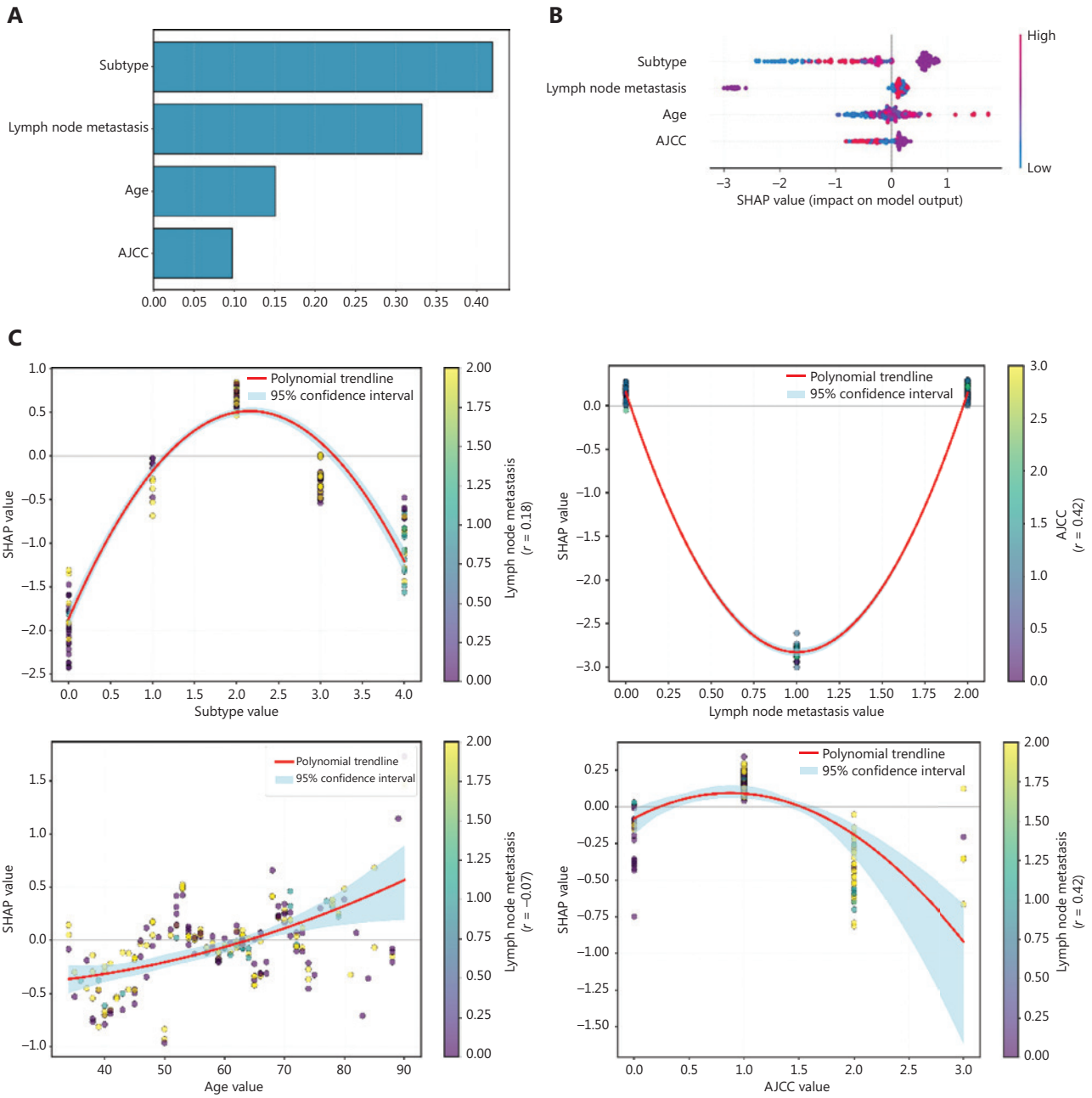


Figure 7 SHAP-based interpretation of the clinical feature model. (A) Global feature importance ranked by mean absolute SHAP values. (B) SHAP beeswarm plot illustrating feature-level contributions. (C) SHAP dependence plots showing interaction patterns for key features. AJCC, American Joint Committee on Cancer; SHAP, SHapley Additive exPlanations.

SHAP-based interpretation of the clinical model

SHAP analysis revealed the relative contributions and directional effects of various clinical features in the prediction model for *PIK3CA* mutation. Molecular subtype exhibited the highest predictive contribution (SHAP importance = 0.4197), and was followed by lymph node metastasis status (SHAP importance = 0.3321), age (SHAP importance = 0.1507), and AJCC stage (SHAP importance = 0.0976) (**Figure 7A**). Further analysis indicated significant differences in the effects of molecular subtypes on mutation risk (**Figure 7B**). Specifically, the luminal A subtype showed the highest tendency toward mutation risk (SHAP value = 1.7702), whereas the basal-like subtype exhibited a lower mutation risk (SHAP value = −0.4977). Additionally, patients without lymph node metastasis demonstrated elevated mutation risk.

SHAP dependence analysis further elucidated the independent influence patterns of each feature (**Figure 7C**). Molecular subtype displayed a clear negative gradient, thus further supporting luminal subtypes as high-risk categories; lymph node metastasis contributed most strongly to prediction at moderate levels; the positive influence of age was weakest in the age range of 40–60 years; and earlier AJCC stages were associated with higher predicted mutation risk. Interfeature correlation analysis indicated a moderate association between AJCC stage and lymph node metastasis in terms of their influence on SHAP values ($r = 0.47$), whereas age, molecular subtype, and other features showed weaker correlations ($r = 0.09$). In summary, the SHAP interpretability framework not only quantified the predictive weight of each feature in the model but also systematically revealed their independent and interactive patterns, thus providing an interpretable computational basis for understanding the clinical predictive pathways of *PIK3CA* mutation.

Discussion

In this study, we developed and validated a multimodal artificial intelligence framework, the MPM, integrating digital histopathology and clinical data to predict *PIK3CA* mutation status in patients with breast cancer. The MPM achieved robust predictive performance across both the internal TCGA validation set and multiple external independent cohorts, with AUC values ranging from 0.680 to 0.745. Importantly, the MPM consistently outperformed each single-modality model (histopathology-only or clinical-only), thereby highlighting the substantial advantage of multimodal fusion in enhancing model generalizability and predictive accuracy. Our work was aimed at developing a reliable histology-based prediction tool for important clinical biomarkers. We systematically evaluated multiple foundation models and algorithms and provided interpretability analyses, thereby providing substantial advances to the field.

Within the histopathology model, we systematically evaluated 7 state-of-the-art pathology foundation models as pretrained feature extractors. H-optimus-0 achieved the best performance on the TCGA dataset (AUC = 0.715), thus indicating its strong capability in capturing histomorphologic patterns associated with *PIK3CA* mutations. Its advanced architecture and pre-training strategies highlighted the notable value of self-supervised learning and contrastive learning in feature extraction from pathological images. Furthermore, the histopathology model's maintenance of favorable generalizability on external datasets (AUC = 0.670–0.680) supported its robustness across multi-center cohorts with heterogeneous staining and scanning conditions.

For the clinical model, the XGBoost model achieved the highest predictive performance (TCGA AUC = 0.694) among multiple machine-learning methods. The strength of XGBoost is its ability to model nonlinear relationships and high-order interactions, which are common in clinical oncology data. On the basis of SHAP interpretability analysis, molecular subtype was identified as the most influential predictor, surpassing all other clinical variables. The luminal A subtype exhibited elevated propensity for mutation risk, in agreement with previously reported associations between these clinical factors and PI3KCA-driven tumor biology in the literature[17,18].

The MPM, which integrates the predicted probabilities from the image and clinical models through logistic regression, achieved an AUC of 0.745 on the TCGA cohort and substantially outperformed the single-modality models. This finding confirmed the complementary nature of digital histopathology and clinical data in predicting *PIK3CA* mutations: the histopathology model captures subtle morphologic patterns within the tumor microenvironment that are associated with underlying genomic alterations, whereas the clinical model incorporates patient-level biological and clinical context. By leveraging both sources of information, the fused model not only improves overall predictive accuracy but also enhances robustness across heterogeneous patient populations. Our results align with the growing trend in medical AI toward using multimodal fusion frameworks to effectively reflect the multifaceted nature of clinical decision-making[19,20]. Importantly, however, the MPM is intended to serve as a pre-screening tool to prioritize cases for confirmatory sequencing but not as a definitive surrogate, given its current absolute performance level.

The interpretability analyses further strengthened the clinical credibility of the developed model. Through attention heatmap visualization, we identified the specific tissue regions that the histopathology model prioritized during prediction. These high-attention areas mapped primarily to regions with dense tumor cell proliferation or distinctive stromal architectures, and demonstrated substantial concordance with pathological patterns known to be associated with *PIK3CA* mutations. Notably, these results aligned with findings from Howard et al. indicating that *PIK3CA*-associated morphologic signatures frequently include enhanced cytoplasmic eosinophilia and elevated tubular or glandular formation[21]. The SHAP analysis further elucidated the contribution and directional effects of each clinical feature on the model's predictions, thereby offering a quantitative basis for understanding the decision-making process of the clinical model. However, existing interpretability methods still fall short of fully elucidating the intricate interactions between morphological and clinical characteristics. Future studies will explore advanced techniques such as concept-based interpretability to improve model transparency and clinical credibility.

Although the MPM demonstrated strong performance across multiple centers, several limitations remain. The relatively small external cohorts might restrict the assessment of generalizability. The clinical model, despite incorporating key variables, might have omitted potentially important factors such as treatment history or family history. In addition,

because the histopathology model might be affected by variations in staining and slide quality, room for further optimization may exist. Moreover, the model achieves optimal performance when both high-quality WSIs and structured clinical data are available, thus potentially posing practical constraints in certain resource-limited clinical settings. The variability in model specificity across multicenter validations is likely to reflect the inherent heterogeneity of real-world clinical and pathological practice. DCA suggested that, across a broad range of threshold probabilities, the model might provide clinical net benefit when used as a decision-support pre-screening tool. However, differences in patient demographics, disease stage distributions, histopathological staining protocols, and digital scanning parameters across centers might contribute to performance variability and also pose practical challenges in large-scale clinical deployment, including infrastructure requirements, data integration, and regulatory considerations. In practice, computationally intensive components of the framework may be deployed *via* centralized or cloud-based infrastructures, to enable flexible integration into clinical workflows without requiring extensive local computing resources. To facilitate more standardized application in multicenter settings, several technical strategies may be considered, such as stain-normalization approaches to mitigate inter-institutional staining variability, privacy-preserving federated learning frameworks to support collaborative modeling across centers, and robust-transfer techniques such as domain adaptation and meta-learning to enhance generalizability to new clinical environments.

Future work will be aimed at expanding external validation cohorts, exploring deeper multimodal fusion strategies, and extending this framework to additional driver-gene mutations and treatment-response prediction. In addition, future studies may explore the incorporation of longitudinal clinical data and treatment history, to enable more dynamic modeling of disease evolution. We also plan to investigate the integration of MPM as a pre-screening decision-support tool within digital pathology workflows, to help prioritize cases for confirmatory molecular testing, thereby strengthening its translational potential in precision oncology.

Conclusions

Herein, we developed a multimodal artificial intelligence framework integrating WSI with structured clinical data to predict *PIK3CA* mutation status in breast cancer. The developed MPM consistently demonstrated robust performance across internal and multi-center external cohorts, and outperformed single-modality approaches, thus highlighting the complementary value of image-derived morphologic features and patient-level clinical information. Interpretability analyses further confirmed that the model captured biologically meaningful patterns associated with *PIK3CA* alterations, thereby strengthening its potential for clinical translation. Collectively, these findings suggested that multimodal fusion is a promising strategy for molecular biomarker prediction that might serve as an effective pre-screening tool to decrease reliance on, rather than replace, routine sequencing assays, thereby optimizing testing strategies and resource allocation to support precision oncology.

Grant support

This study was financially supported by the Hebei Natural Science Foundation (Grant No. H2024206504), the Medical Science Research Project of Hebei (Grant No. 20260484, 20260530), and the Fundamental Research Funds for the Central Universities (Grant No. 20822041J4123).

Conflict of interest statement

No potential conflicts of interest are disclosed.

Author contributions

Conceived and designed the analysis: Yueping Liu, Xiao Han, Lianghong Teng.
Collected the data: Jiaxian Miao, Qi Liu, Jianing Zhao, Shishun Fan, Shenwen Wang, Si Wu, Jinze Li, Huirui Zhang, Meng Zhang.
Contributed data or analysis tools: Feng Ye, Hong Bu.
Performed the analysis: Jiaxian Miao, Jianing Zhao.
Wrote the article: Jiaxian Miao, Jianing Zhao.

Data availability statement

The data generated in this study are publicly available in The Cancer Genome Atlas (TCGA) at https://portal.gdc.cancer.gov/. The multicenter external validation data are available on reasonable request from the corresponding author.

References

1. Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med. 2025; 31: 1154-62.
2. Giaquinto AN, Sung H, Newman LA, Freedman RA, Smith RA, Star J, et al. Breast cancer statistics 2024. CA Cancer J Clin. 2024; 74: 477-95.
3. Xiong X, Zheng LW, Ding Y, Chen YF, Cai YW, Wang LP, et al. Breast cancer: pathogenesis and treatments. Signal Transduct Target Ther. 2025; 10: 49.
4. Miricescu D, Totan A, Stanescu-Spinu II, Badoiu SC, Stefani C, Greabu M. PI3K/AKT/mTOR signaling pathway in breast cancer:

from molecular landscape to clinical aspects. Int J Mol Sci. 2020; 22: 173.
5. Xiao W, Zhang G, Chen B, Chen X, Wen L, Lai J, et al. Mutational landscape of PI3K-AKT-mTOR pathway in breast cancer: implications for targeted therapeutics. J Cancer. 2021; 12: 4408-17.
6. Jiang YZ, Ma D, Jin X, Xiao Y, Yu Y, Shi J, et al. Integrated multiomic profiling of breast cancer in the Chinese population reveals patient stratification and therapeutic vulnerabilities. Nat Cancer. 2024; 5: 673-90.
7. Kianfar E, Alrudainy AM, Al-Zaalan AR, Kadhum WR, Jassim AY. PI3K inhibitors: efficacy in diverse cancer forms. Cancer Treat Res Commun. 2025; 45: 101028.
8. Nadjkovic K, Lonabaugh K. Alpelisib: a novel agent for PIK3CA-related overgrowth spectrum. J Pediatr Pharmacol Ther. 2023; 28: 590-4.
9. Alam MR, Seo KJ, Abdul-Ghafar J, Yim K, Lee SH, Jang HJ, et al. Recent application of artificial intelligence on histopathologic image-based prediction of gene mutation in solid cancers. Brief Bioinform. 2023; 24: bbad151.
10. Giuliano AE, Edge SB, Hortobagyi GN. Eighth edition of the AJCC Cancer Staging Manual: breast cancer. Ann Surg Oncol. 2018; 25: 1783-5.
11. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng. 2021; 5: 555-70.
12. 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: 850-62.
13. Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. Med Image Anal. 2022; 81: 102559.
14. Zimmermann E, Vorontsov E, Viret J, Casson A, Zelechowski M, Shaikovski G, et al. Virchow2: scaling self-supervised mixed magnification models in pathology. 2024. Preprint at https://doi.org/10.48550/arXiv.2408.00738.
15. Saillard C, Jenatton R, Llinares-López F, Mariet Z, Cahané D, Durand E, et al. H-optimus-0 [Internet]. 2024. Available from: https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0.
16. Lu MY, Chen B, Williamson DFK, Chen RJ, Liang I, Ding T, et al. A visual-language foundation model for computational pathology. Nat Med. 2024; 30: 863-74.
17. Mosele F, Stefanovska B, Lusque A, Tran Dien A, Garberis I, Droin N, et al. Outcome and molecular landscape of patients with *PIK3CA*-mutated metastatic breast cancer. Ann Oncol. 2020; 31: 377-86.
18. Martínez-Sáez O, Chic N, Pascual T, Adamo B, Vidal M, González-Farré B, et al. Frequency and spectrum of *PIK3CA* somatic mutations in breast cancer. Breast Cancer Res. 2020; 22: 45.
19. Niehues JM, Quirke P, West NP, Grabsch HI, van Treeck M, Schirris Y, et al. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: a retrospective multi-centric study. Cell Rep Med. 2023; 4: 100980.
20. Ding T, Wagner SJ, Song AH, Chen RJ, Lu MY, Zhang A, et al. A multimodal whole-slide foundation model for pathology. Nat Med. 2025; 31: 3749-61.
21. Howard FM, Hieromnimon HM, Ramesh S, Dolezal J, Kochanny S, Zhang Q, et al. Generative adversarial networks accurately reconstruct pan-cancer histology from pathologic, genomic, and radiographic latent features. Sci Adv. 2024; 10: eadq0856.

Cite this article as: Miao J, Liu Q, Zhao J, Fan S, Wang S, Ye F, et al. Multimodal artificial intelligence predicts *PIK3CA* mutation in breast cancer from digital pathology and clinical data: a multicenter study. Cancer Biol Med. 2026; 23: 430-450. doi: 10.20892/j.issn.2095-3941.2025.0771