Breast Cancer Research

RESEARCH Open Access

Multimodal recurrence risk prediction model for HR+/HER2- early breast cancer following adjuvant chemo-endocrine therapy: integrating pathology image and clinicalpathological features

Xiaoyan Wu[1,2†], Yiman Li[3†], Jilong Chen[3], Jie Chen[2], Wenchuan Zhang[1,2], Xunxi Lu[1,2], Xiaorong Zhong[4], Min Zhu[3], Yuhao Yi[2,3*] and Hong Bu[1,2]

Abstract

Background In HR+/HER2- early breast cancer (EBC) patients, approximately one-third of stage II and 50% of stage III patients experience recurrence, with poor outcomes after recurrence. Given that these patients commonly undergo adjuvant chemo-endocrine therapy (C-ET), accurately predicting the recurrence risk is crucial for optimizing treatment strategies and improving patient outcomes.

Methods We collected postoperative histopathological slides from 1095 HR+/HER2- EBC who received C-ET and were followed for more than five years at West China Hospital, Sichuan University. Two deep learning pipelines were developed and validated: ACMIL-based and CLAM-based. Both pipelines, designed to predict recurrence risk post-treatment, were based on pretrained feature encoders and multi-instance learning with attention mechanisms. Model performance was evaluated using a five-fold cross-validation approach and externally validated on HR+/HER2- EBC patients from the TCGA cohort.

Results Both ACMIL-based and CLAM-based pipelines performed well in predicting recurrence risk, with UNI-ACMIL demonstrating superior performance across multiple metrics. The average area under the curve (AUC) for the UNI-ACMIL pipeline in the five-fold cross-validation test set was 0.86 ± 0.02, and 0.80 ± 0.04 in the TCGA cohort. In the five-fold cross-validation test sets, effectively stratified patients into high-risk and low-risk groups, demonstrating significant prognostic differences. Hazard ratios for recurrence-free survival (RFS) ranged from 5.32 (95% CI 1.86-15.12) to 15.16 (95% CI 3.61-63.56). Moreover, among six different multimodal recurrence risk models, the WSI-based risk score was identified as the most significant contributor.

[†] Xiaoyan Wu and Yiman Li contributed equally to this work.

*Correspondence:
Yuhao Yi
yuhaoyi@scu.edu.cn

Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusion Our multimodal recurrence risk prediction model is a practical and reliable tool that enhances the predictive power of existing systems relying solely on clinicopathological parameters. It offers improved recurrence risk prediction for HR+/HER2- EBC patients following adjuvant C-ET, supporting personalized treatment and better patient outcomes.

Keywords HR+/HER2- early breast cancer, Recurrence risk, Adjuvant chemo-endocrine therapy, Deep learning pipelines, Pathology image

Introduction

HR+/HER2- breast cancer is the most common subtype, accounting for approximately 70%~75% of all breast cancer cases, with the majority diagnosed at early stages (stage I-III). Despite generally favorable prognoses, patients with HR+/HER2- early breast cancer (EBC) remain at risk for recurrence. Studies show that 27%~37% of stage II patients and 46%~57% of stage III patients experience recurrence after completing five years of standard endocrine therapy (ET) [1], while approximately 20% relapse following adjuvant chemo-endocrine therapy (C-ET) [2]. These figures suggest the presence of a more aggressive subgroup within HR+/HER2- EBC, highlighting the importance of accurately identifying these patients. This is crucial for tailoring treatment plans and formulating appropriate recommendations for adjuvant therapies aimed at preventing recurrence. Common prognostic factors in HR+/HER2- breast cancer include tumor size, lymph node status, Ki-67 index, lymphovascular invasion, histological grade, and multigene assay scores. Risk prediction models incorporating these factors have shown significant prognostic value. However, these models primarily focus on the risk of recurrence after endocrine therapy, often overlooking the impact of chemotherapy and missing additional valuable information contained in histopathological slides.

Histopathological slides of breast cancer contain a wealth of prognostically relevant information, including cellular atypia, duct formation, mitotic activity and vascular tumor emboli. However, inter-observer and intra-observer variability often arise when pathologists visually quantify these features. To address this, several studies have proposed using deep learning to automate quantification in order to improve assessment consistency [3–5]. In recent years, deep learning has demonstrated exceptional performance in image interpretation tasks, enabling automatic feature extraction without the need for manually predefined structures of interest [6–8]. A key advantage of deep learning-based histopathological prediction is that these models are not constrained by prior knowledge of predefined image features. Instead, they can evaluate any histopathological pattern and incorporate it alongside other coexisting patterns, thereby generating risk scores that reflect a comprehensive assessment of the tissue.

To predict the recurrence risk in HR+/HER2- EBC patients following adjuvant C-ET, we developed two deep learning pipelines based on pretrained tissue-specific feature extractors and multi-instance learning with attention mechanisms. Additionally, we integrated the WSI-based risk score with clinicopathological factors to construct a multimodal recurrence risk prediction model. The model was validated using five-fold cross-validation method, and we explored the pathophysiological mechanisms associated with the WSI-based risk score to provide biological interpretability for the deep learning-based prediction.

Currently, no established model exists for predicting the recurrence risk of HR+/HER2- EBC patients following adjuvant C-ET. Therefore, in this study, we combined histopathological slide information with clinicopathological factors to construct a multimodal model capable of predicting recurrence risk in HR+/HER2- EBC patients. This approach enables the identification of high-risk patients who may still experience recurrence, ultimately aiding in the formulation of precise therapeutic strategies.

Materials and methods

Patient cohort

In this retrospective study, we utilized H&E-stained slides of HR+/HER2- EBC from the Pathology Department of West China Hospital, Sichuan University, for model training and validation. As of May 31, 2023, a total of 30,004 breast cancer patients were retrieved from the Breast Cancer Management Information System at West China Hospital (WCH). Medical records, pathological diagnoses, and treatment information were collected by professional physicians. Patients were followed up through outpatient visits or telephone calls every 3-4 months during the first 2 years after initial diagnosis, every 6 months for the subsequent 3-5 years, and annually thereafter. The inclusion criteria for the study were as follows: (1) unilateral primary invasive breast cancer, clinically staged as I-III at the time of initial diagnosis; (2) patients who received adjuvant C-ET within 3 months postoperatively, without any preoperative treatments; (3) patients with a clear postoperative pathological diagnosis. The exclusion criteria were: (1) patients with clinical stage IV; (2) patients with multifocal or bilateral invasive breast cancer; (3) patients who received only adjuvant ET or CT; (4) patients with incomplete clinical information. Detailed inclusion and exclusion criteria are provided in the Figure S1. We also used an external validation cohort consisting of 325 HR+/HER2- EBC patients from The Cancer Genome Atlas (TCGA) database, who met the inclusion criteria, to evaluate the deep learning models.

Experimental design

The deep learning models generated patient-level risk scores and assessed the model's ability to predict recurrence risk in HR+/HER2- EBC patients following adjuvant C-ET. Model performance was evaluated using area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1 score. Patients were classified into high-risk (above the threshold) and low-risk (below or equal to the threshold) groups based on the median WSI-based risk score derived from the training set. In the WCH cohort, we performed the following analyses: First, we assessed whether the WSI-based risk score provided additional prognostic value across distinct pathological and clinical grades. Kaplan-Meier analysis and log-rank tests were used to compare differences in recurrence survival between the groups. Subsequently, we applied the Cox proportional hazards model to evaluate recurrence survival differences within these groups. Next, we integrated the WSI-based risk score with clinicopathological factors to develop multimodal recurrence risk prediction models. This integration was performed using classical Cox proportional hazards regression (CPH), elastic net Cox (EN-Cox), gradient boosting regression tree (GBRT), Lasso-Cox, Ridge-Cox, and random survival forest (RSF). The training labels for all machine learning models included both the recurrence status of the patients, with detailed parameters provided in Table S1. We then performed 5-fold cross-validation, with data divisions corresponding to those used in the deep learning model, to compare the performance of these different models. The primary analysis endpoints of this study was recurrence-free survival (RFS) and the secondary analysis endpoint was overall survival (OS). RFS was defined as the interval from surgery to recurrence, metastasis, death from breast disease, or last follow-up. OS was defined as the time from surgery to death from any cause or the last follow-up date.

Statistics

Statistical analyses were performed using R software version 4.3.2. The survcomp package was used to calculate the concordance index (C-index) and its 95% confidence interval (CI). Time-dependent receiver operating characteristic (ROC) analysis was conducted using the timeROC package. Kaplan-Meier analysis and log-rank tests were carried out using the survival and survminer packages. Multivariate analysis was conducted using the Cox proportional hazards model from the survival package. All survival models (CPH, EN-Cox, GBRT, Lasso-Cox, Ridge-Cox, and RSF) were constructed using the scikit-survival package in Python. All the statistical tests were two-sided, with a $P < 0.05$ considered to indicate statistical significance.

Image preprocessing

H&E-stained 4-μm FFPE tissue sections from 1095 HR+/HER2- EBC patients in the WCH cohort were used to generate at least one representative tumor section for per patient, resulting in a total of 1304 slides. These slides were scanned at 40× magnification using a UNIC digital pathology scanner (PRECICE 600 series). Tumor regions of interest (ROIs) on all WSIs were manually annotated by a professional pathologist using ASAP software version 1.8 (available at https://github.com/computationalpathologygroup/ASAP/releases). All WSIs from the WCH and TCGA cohorts were segmented into non-overlapping tiles of varying sizes to meet the requirements of the respective feature encoders. Specifically, CTransPath utilized tiles at 10× magnification with a resolution of 224 × 224 pixels, CONCH required tiles at 20× magnification with a resolution of 448 × 448 pixels, REMEDIS employed tiles at 20× magnification with a resolution of 224 × 224 pixels, and UNI used tiles at 20× magnification with a resolution of 256 × 256 pixels.

Deep learning methods

We developed two pipelines, ACMIL-based and CLAM-based, for training and validating deep learning models. The recurrence status of patients was used as the training label for the deep learning models to derive recurrence risk scores for HR+/HER2- EBC patients. Both pipelines employed pretrained tissue-specific encoders, including CTransPath (an SSL algorithm based on MoCo v3) [9], CONCH (a vision-language foundation model pretrained on diverse tissue pathology images and biomedical texts) [10], REMEDIS (combining large-scale supervised transfer learning with natural images and intermediate contrastive self-supervised learning for medical images) [11], UNI (an SSL algorithm based on DINOv2) [12], and Virchow (also an SSL algorithm based on DINOv2) [13]. These encoders transformed tiles into feature vectors, which were then aggregated into a bag for each WSI. Patient-level recurrence risk scores were predicted using multiple-branch attention mechanism-based ACMIL [14] and gated attention mechanism-based CLAM [15], ultimately generating risk score outputs. All experiments were conducted at the patient-level, utilizing a five-fold cross-validation approach. The WCH cohort was randomly into training, validation and test sets with a final patient-level ratio of 7:1:2. The TCGA cohort served as an external validation set. Given that the data included manually annotated tumor regions, four deep learning models (ACMIL-based, Roi-ACMIL-based, CLAM-based and Roi-CLAM-based) were trained in this study.

Visualization and interpretability

To gain insights into the WSI-based risk scores and the internal predictive patterns of the models, we visualized the spatial distribution of attention and predictive scores using heatmaps. These heatmaps highlighted the tiles (top tiles) with the highest attention-weighted predictive scores. Attention heatmaps were generated for all patients in the test set to reveal the regions focused on by the models during prediction. Additionally, we analyzed the underlying features of the top 20 tiles. For deeper insights, we also provided weight-score heatmaps and identified top tiles for the four most representative patients with the highest risk scores as determined by deep learning model. Attention heatmaps visualize the image regions focused on by the model during prediction, while weight-score heatmaps display the relative importance of different areas within the image. Top tiles represent the specific regions with the most significant influence on the model's predictive performance. To explore the biological characteristics associated with the WSI-based risk score, transcriptomic sequencing data from 99 mRNA-sequencing samples in the WCH cohort were analyzed. Differentially expressed genes (DEGs) between high-risk and low-risk groups were identified using the edgeR package in R, followed by gene set enrichment analysis (GSEA) on these DEGs. Furthermore, we assessed the relationship between the WSI-based risk score and immune infiltration by calculating tumor-infiltrating immune cell scores using the CIBERSORT algorithm.

Results

Development and performance evaluation of deep learning pipelines based on attention mechanisms

We developed two deep learning pipelines: CTP-ACMIL, which combines CTransPath as the feature encoder with ACMIL for downstream feature aggregation, and CTP-CLAM, which integrates CTransPath with the CLAM feature aggregation algorithm. The ACMIL-based pipeline is illustrated in Fig. 1. For model development, we performed five-fold cross-validation on data from 1,095 HR+/HER2- EBC patients. The final dataset included 1034 WSIs, as some patients had multiple slides, with complete clinical and pathological information available for only 968 patients. The clinicopathological characteristics of all patients are summarized in Table 1, with their distribution across folds (fold 0-fold 4) detailed in Table 1, Table S2 and Table S3. Although recent advancements in deep learning have enabled tasks without the need for manual ROI annotations, this study

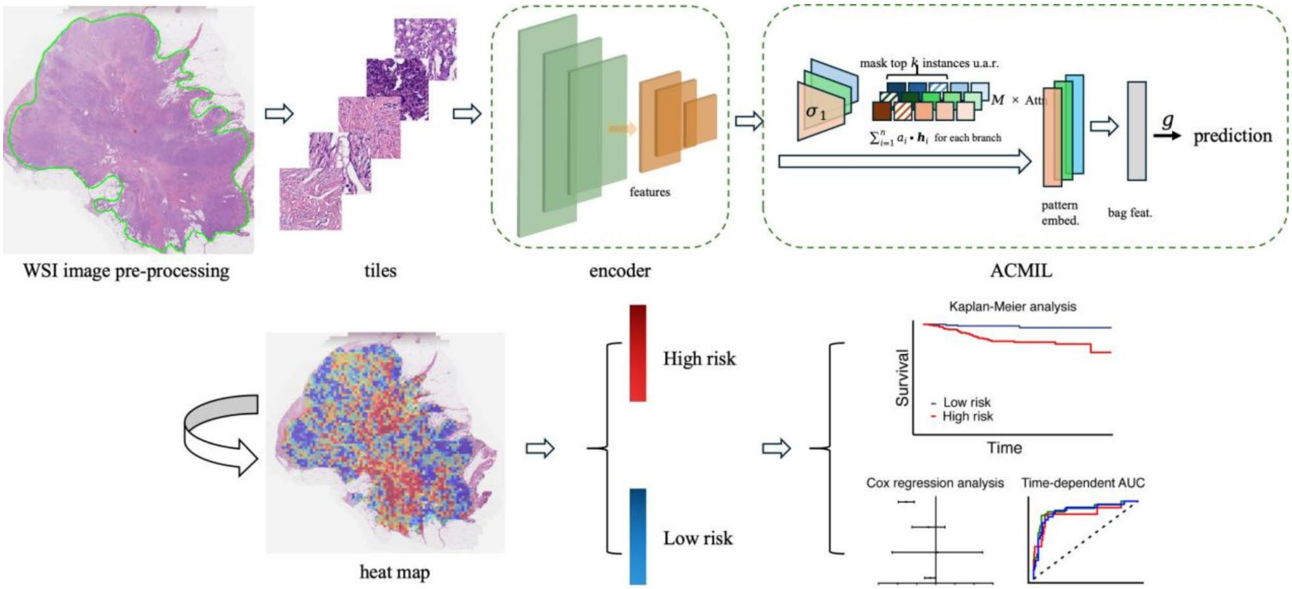


Fig. 1 Deep learning pipeline for predicting recurrence risk following adjuvant C-ET in HR+/HER2- EBC. (**A**) WSIs with manually annotated tumor regions were divided into nonoverlapping 224 × 224 pixel tiles at 10× magnification. (**B**) Features were extracted from each tile using a self-supervised feature encoder CTransPath, resulting in 768-dimensional vector features. (**C**) Patient-level recurrence risk prediction was achieved by aggregating all feature vectors from each WSI into a bag using the multiple branch attention mechanism-based ACMIL, resulting in the generation of the final risk score output. (**D**) The model was trained to classify prognostic outcomes for WSIs, assigning attention scores to each tile. Attention heatmaps showed the attention scores assigned by the model to each tile for predicting patient recurrence, with blue indicating low attention and red indicating high attention. (**E**) Patients were categorized into high risk and low risk groups based on the median risk score from the training set, which served as the threshold for classification. These groups were then utilized for subsequent survival analysis. C-ET, chemo-endocrine therapy; EBC, early breast cancer; WSIs, whole slide images

Table 1 Clinicopathological characteristics of patients and the distribution of characteristics across the training, validation, and test sets in fold 0

Characteristic	*N* = 968	Fold 0		
		Training *N* = 660	Validation *N* = 111	Test *N* = 197
Age				
Mean ± SD	49 ± 9	49 ± 9	49 ± 9	48 ± 9
Tumor size				
≤ 2 cm	414 (43%)	277 (42%)	57 (51%)	80 (41%)
> 2 cm	554 (57%)	383 (58%)	54 (49%)	117 (59%)
LN				
0	410 (42%)	286 (43%)	40 (36%)	84 (43%)
1 ~ 3	364 (38%)	243 (37%)	43 (39%)	78 (40%)
≥ 4	194 (20%)	131 (20%)	28 (25%)	35 (18%)
Clinical stage				
I	203 (21%)	145 (22%)	22 (20%)	36 (18%)
II	564 (58%)	379 (57%)	60 (54%)	125 (63%)
III	201 (21%)	136 (21%)	29 (26%)	36 (18%)
Grade				
1	39 (4.0%)	25 (3.8%)	3 (2.7%)	11 (5.6%)
2	554 (57%)	390 (59%)	54 (49%)	110 (56%)
3	375 (39%)	245 (37%)	54 (49%)	76 (39%)
LVI				
0	886 (92%)	609 (92%)	100 (90%)	177 (90%)
1	82 (8.5%)	51 (7.7%)	11 (9.9%)	20 (10%)
ER				
Mean ± SD	0.82 ± 0.17	0.83 ± 0.16	0.81 ± 0.19	0.81 ± 0.18
PR				
Mean ± SD	0.60 ± 0.32	0.61 ± 0.32	0.56 ± 0.34	0.60 ± 0.33
HER2				
0	263 (27%)	172 (26%)	33 (30%)	58 (29%)
1	705 (73%)	488 (74%)	78 (70%)	139 (71%)
Ki67				
Mean ± SD	0.26 ± 0.17	0.26 ± 0.18	0.26 ± 0.16	0.28 ± 0.17
Molecular subtype				
0	209 (22%)	159 (24%)	23 (21%)	27 (14%)
1	759 (78%)	501 (76%)	88 (79%)	170 (86%)
RFS status				
No recurrence	838 (87%)	579 (88%)	90 (81%)	169 (86%)
Recurrence	130 (13%)	81 (12%)	21 (19%)	28 (14%)

LN, lymph node; LVI: lymphatic vessel infiltration; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor; RFS, recurrence-free survival

compared model performance with and without manually annotated ROIs. Both models accurately predicted recurrence risk in HR+/HER2- EBC from unannotated WSIs, achieving high performance. The AUC values in the five-fold cross-validation test sets were 0.81 ± 0.02 for CTP-ACMIL and 0.80 ± 0.02 for CTP-CLAM. When incorporating manually annotated ROIs, the average AUCs were 0.82 ± 0.02 for RoiCTP-ACMIL and 0.78 ± 0.03 for RoiCTP-CLAM. Additional performance metrics, including accuracy, precision, recall, and F1 score showed similar trends to the AUC values (Fig. 2A).

A comprehensive assessment of the model performance metrics revealed that RoiCTP-ACMIL demonstrated superior predictive performance (Fig. 2A). Consequently, we selected the model trained on the ROI-containing WSI dataset using the ACMIL algorithm as our final recurrence risk prediction model. The RoiCTP-ACMIL model achieved AUC values of 0.84 (95% CI: 0.77-0.84), 0.81 (95% CI: 0.77-0.84), 0.80 (95% CI: 0.77-0.84), 0.81 (95% CI: 0.77-0.84), and 0.85 (95% CI: 0.77-0.84) across the five-fold cross-validation test sets (Fig. 2C). Detailed AUC values with corresponding 95% confidence intervals for the validation sets are also presented in Fig. 2B. WSI-based risk scores were generated for all 1,095 HR+/HER2- EBC patients using the test sets from the five-fold cross-validation process. An analysis of the distribution of the WSI-based risk scores and recurrence status indicated that patients with higher scores generally had a higher recurrence rate compared to those with lower scores (Fig. 2D). Patients were further stratified into high- and low-risk groups based on the median WSI-based risk score derived from the training set. Kaplan-Meier survival analysis revealed that the high-risk group had significantly poorer RFS compared to the low-risk group in both test and validation sets across the five-fold cross-validation (Fig. 2E-F). Similarly, OS was worse in the high-risk group compared to the low-risk group (Figure S2).

Impact of different feature encoders on the performance of deep learning models

To further evaluate the impact of different feature encoders on the performance of deep learning models, we compared the performance of five feature encoders—CTransPath, UNI, CONCH, Virchow, and REMEDIS—within the ACMIL and CLAM models. Model performance was assessed using four metrics: AUC, accuracy, F1-score, and recall. In the five-fold cross-validation results on the test of the WCH cohort, all five feature encoders exhibited satisfactory performance in both the ACMIL and CLAM models, with AUC values ranging from 0.75 ± 0.08 to 0.86 ± 0.02 (Fig. 3A). Further analysis revealed that, except for Virchow, the remaining four feature encoders exhibited a consistent trend of better performance in ACMIL compared to CLAM, as indicated by the AUC metric. Notably, UNI showed the most significant improvement, with ACMIL outperforming CLAM by up to 7%. Similarly, for the accuracy metric, ACMIL consistently outperformed CLAM across all feature encoders, with UNI achieving the largest improvement of 13%. Additionally, within the ACMIL model, the UNI encoder exhibited superior performance across multiple evaluation metrics compared to the other encoders (Fig. 3A). The average C-index values for the UNI encoder were 0.828 ± 0.006 in the training

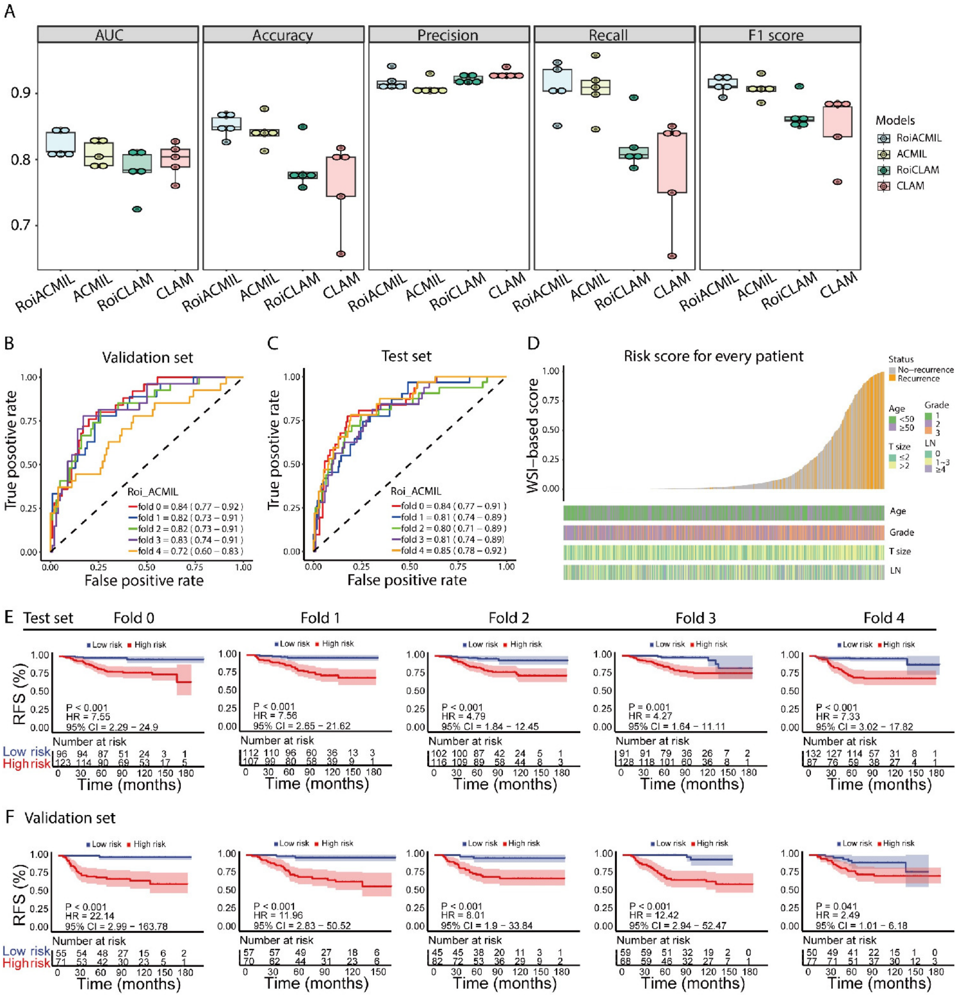


Fig. 2 Evaluation of deep learning model performance and Kaplan-Meier analysis of WSI-based risk score. (**A**). Performance metrics (AUC, accuracy, precision, recall, and F1 score) of the four deep learning models: CTP-ACMIL, RoiCTP-ACMIL, CTP-CLAM and RoiCTP-CLAM. (**B-C**). ROC curves for recurrence risk score of the RoiCTP-ACMIL model in 5-fold cross-validation. (**D**). Heatmap distribution of WSI-based risk score and patient recurrence status for the RoiCTP-ACMIL model. (**E-F**). Kaplan-Meier curves for RFS in the validation set and test set. HR and 95% CI were calculated using the Cox proportional hazards model. P values were calculated using the log-rank test. WSI, whole slide image; RFS, recurrence-free survival

set, 0.818 ± 0.015 in the validation set, and 0.837 ± 0.023 in the test set (Table S4). Model performance was also evaluated on the TCGA cohort. The results showed that CTransPath demonstrated poor performance across all metrics in both the ACMIL and CLAM models, with AUC values of 0.54 ± 0.03 and 0.54 ± 0.04, respectively. In contrast, the other four feature encoders significantly outperformed CTransPath, achieving AUC values as high as 0.84 ± 0.03 (Fig. 3B).

In the heatmaps of WSI-based risk scores generated by the UNI feature encoder, we observed a pattern similar to that of the CTransPath feature encoder: patients with higher scores generally exhibited higher recurrence rates compared to those with lower scores (Fig. 3C). Additionally, WSI-based risk scores derived from the UNI feature encoder demonstrated a higher hazard ratio (HR) for risk stratification of RFS in HR+/HER2- EBC patients compared to those derived from the CTransPath feature encoder across five-fold cross-validation. The HR values for each fold were as follows (UNI vs. CTransPath): fold 0: 9.73 vs. 7.55; fold 1: 15.16 vs. 7.56; fold 2: 5.32 vs. 4.79; fold 3: 12.88 vs. 4.27; and fold 4: 8.52 vs. 7.33 (Figs. 2E and 3D-H).

Additional survival analyses were conducted for patient subgroups defined by distinct clinicopathological variables using the WSI-based risk score. In line with recent recommendations to administer CDK4/6 inhibitors to high-risk HR+/HER2- EBC patients, we re-stratified patients from cohorts 1 and 2 of the monarchE study. Cohort 1 included patients with ≥ 4 positive axillary lymph nodes or 1-3 positive axillary lymph nodes with a tumor size ≥ 5 cm or histological grade 3, while cohort 2 included patients with 1-3 positive axillary lymph nodes, tumor size < 5 cm, histological grade < 3, and a Ki67

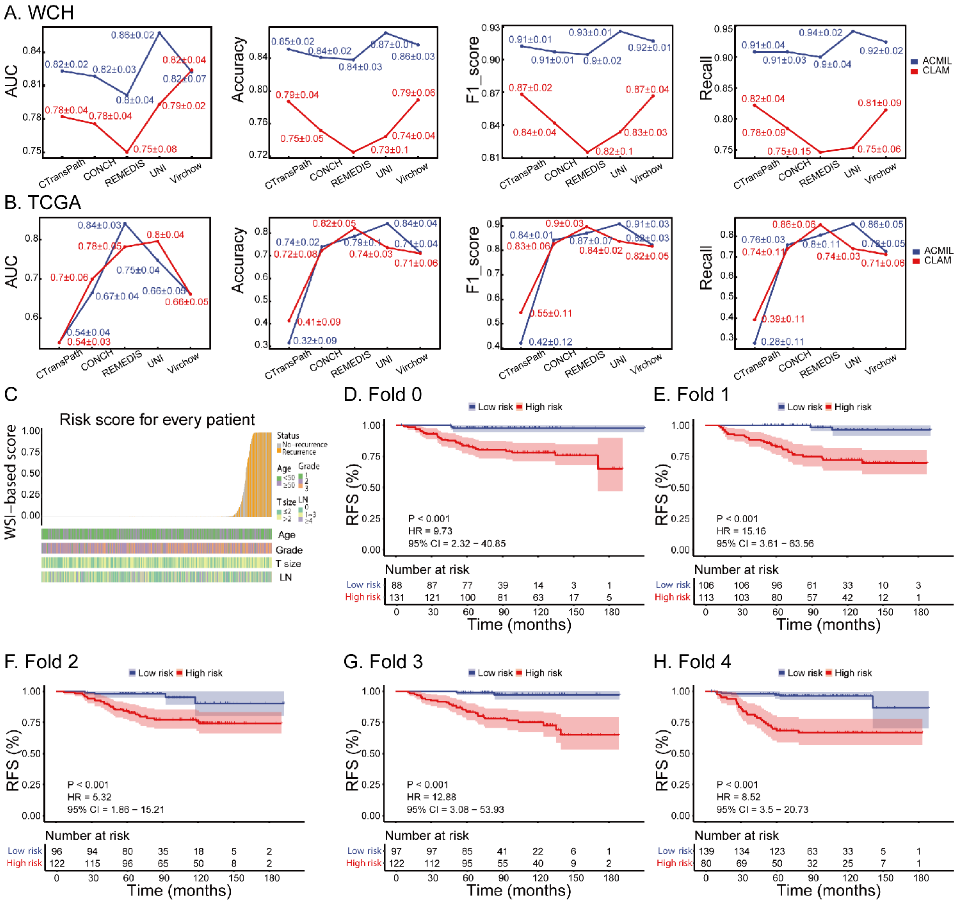


Fig. 3 Comparison of the impact of different feature encoders on deep learning model performance and Kaplan-Meier analysis of WSI-based risk score. (**A-B**). Performance metrics (AUC, accuracy, recall, and F1 score) of five feature encoders (CTransPath, UNI, CONCH, Virchow, and REMEDIS) in the ACMIL-Based deep learning model for the WCH and TCGA cohorts. (**C**). Heatmap distribution of WSI-based risk score and patient recurrence status for the UNI-ACMIL model. (**D-H**). Kaplan-Meier curves for RFS in the test set. HR and 95% CI were calculated using the Cox proportional hazards model. P values were calculated using the log-rank test. WSI, whole slide image; RFS, recurrence-free survival

index ≥ 20%. Subgroup analyses of the test set, stratified by pathological variables (e.g., Ki67 index, histological grade, and molecular subtype) and clinical variables (e.g., staging, risk stratification, and clinical trial cohorts), demonstrated that WSI-based risk scores effectively facilitated risk re-stratification within these subgroups (Figure S3-S7).

Interpretability analysis of deep learning model

We generated attention weight-score heatmaps and selected the top 20 tiles for all patients in the test set to analyze their underlying features. Upon examination by a professional pathologist, the top tiles in the high-risk group were predominantly characterized by cord-like or sheet-like arrangements of tumor cells with large nuclei, prominent nucleoli, and vacuolated or hyperchromatic nuclei appearances. Additionally, high-density stromal regions with large nuclei and spindle-shaped cells, along with areas exhibiting minimal lymphocyte infiltration, were identified as key areas of focus. Figure 4 presents four representative heatmaps along with their corresponding tiles, derived from the patients with the highest WSI-based risk scores in the test set of the deep learning model. For each patient, six of the most representative tiles were selected from the top 20 tiles ranked by attention weights. These high-attention tiles highlight the model's capability to identify high-risk regions. Bioinformatics analysis of transcriptomic data from 99 patients in the WCH cohort identified 150 DEGs between the high-risk and low-risk groups (Figure S8A). GSEA revealed that these DEGs were significantly enriched in four pathways: IFN-α response, IFN-γ response, allograft rejection, and KRAS signaling (down regulated) (Figure S8B). Additionally, CIBERSORT analysis indicated that the high-risk group exhibited lower proportions of T cells gamma delta ($P < 0.01$) compared to the low-risk group, while monocytes were more abundant in the high-risk group ($P = 0.02$) (Figure S8C-D). No statistically significant differences were observed in other immune-related cells types between the two groups.

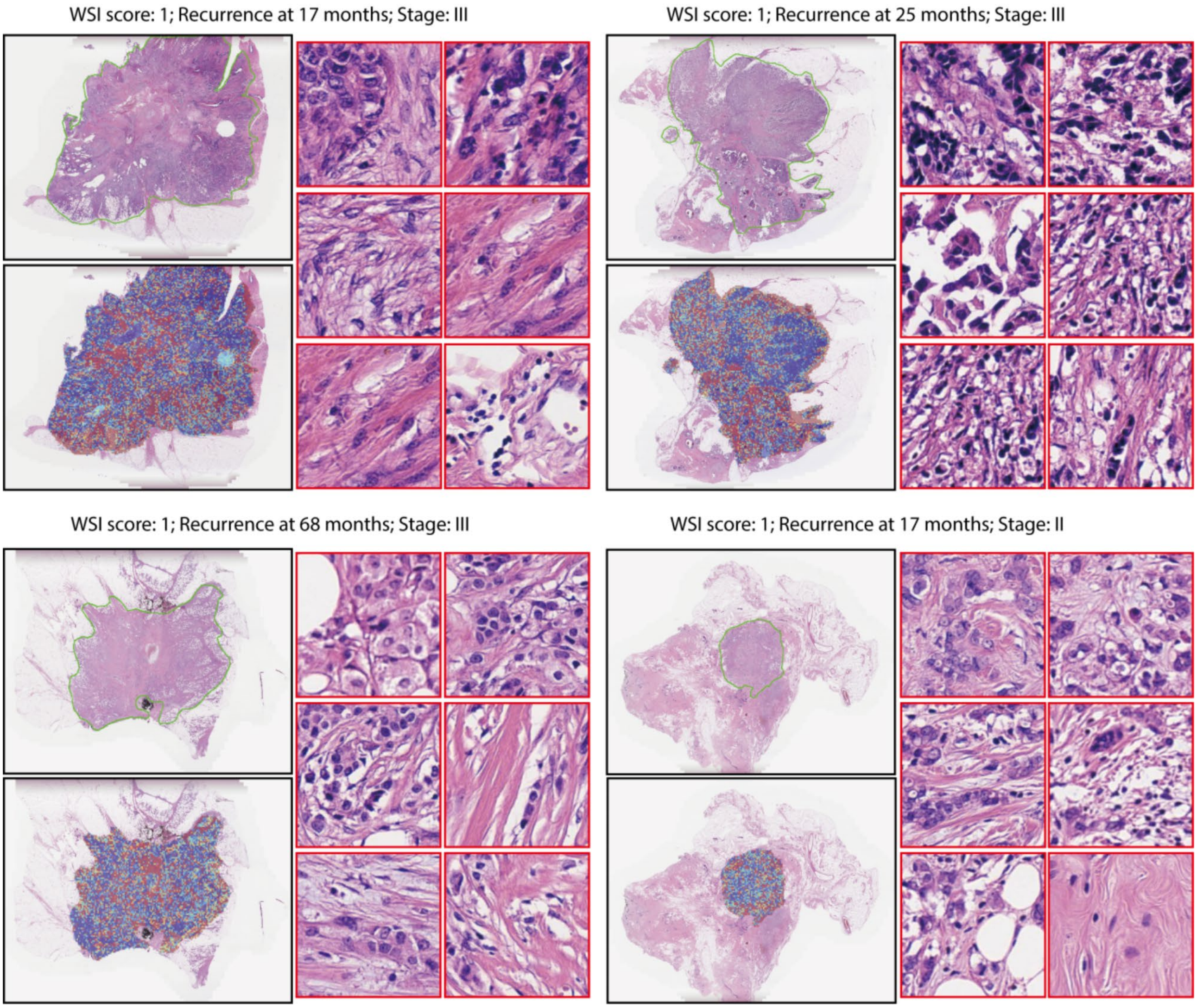
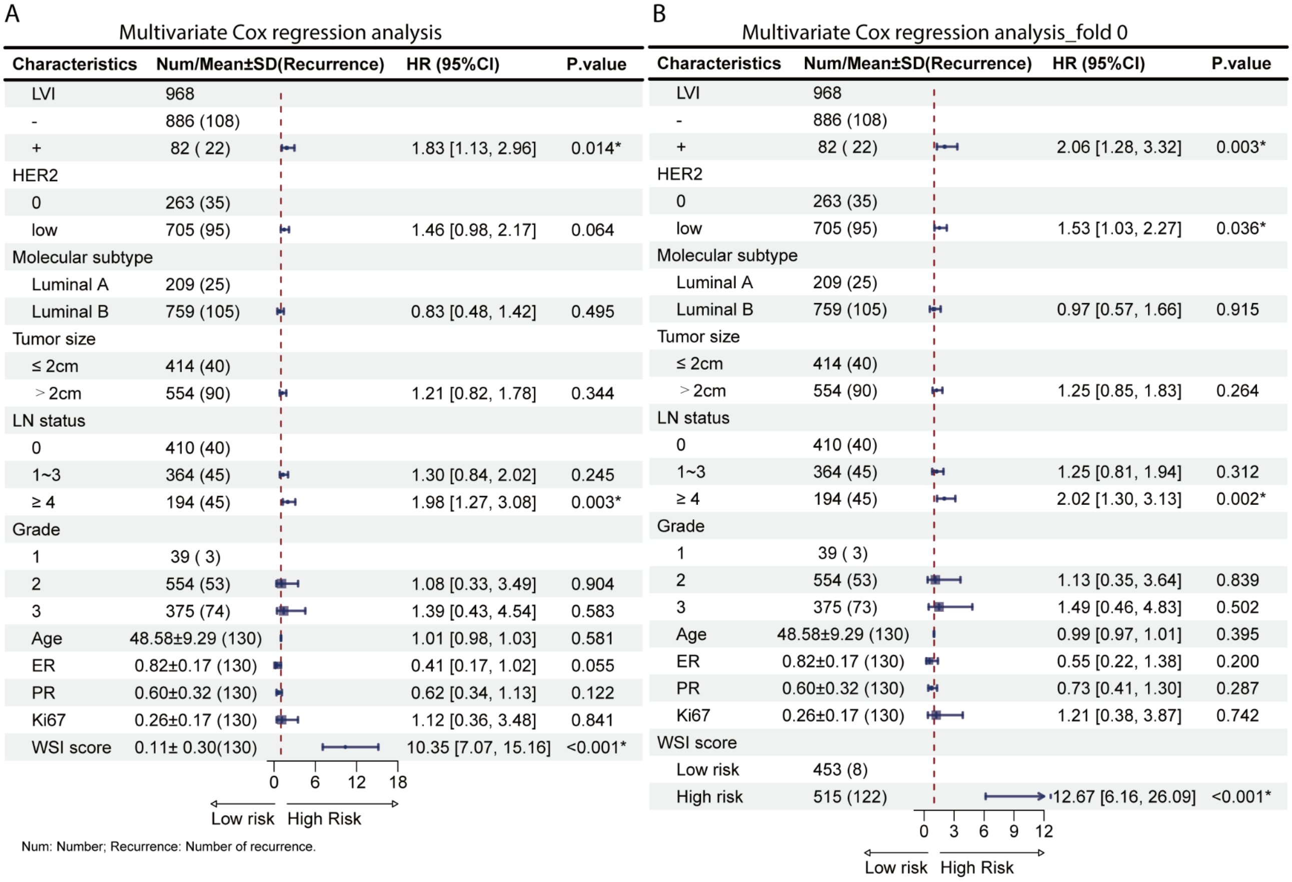

WSI score: 1; Recurrence at 17 months; Stage: III
WSI score: 1; Recurrence at 25 months; Stage: III
WSI score: 1; Recurrence at 68 months; Stage: III
WSI score: 1; Recurrence at 17 months; Stage: II

Fig. 4 Visualization of the top 4 representative cases with the highest risk scores based on the UNI-ACMIL model. In each case, images with black borders display the images of the manually annotated tumor regions (upper left), and the attention weight score heatmaps (lower left). The colors of the heatmaps correspond to the attention scores of tiles within the WSI, with red indicating tiles that have a significant impact on the model's prediction and blue indicating tiles with a smaller impact. The images with red borders display six representative tiles from the top 20 tiles. WSI, whole slide image

Model integration of WSI-based risk score and clinicopathological features

To evaluate whether the WSI-based risk score serves as an independent prognostic factor for predicting RFS in HR+/HER2- EBC, we performed a multivariable Cox regression analysis. This analysis included 968 patients with complete clinicopathological data, adjusting for relevant clinicopathological variables. The results confirmed that the WSI-based risk score serves as an independent prognostic factor, whether analyzed as a categorical or continuous variable, across five-fold cross-validation (Fig. 5, Figure S9). Building on these results, we developed an 11-dimensional feature set by integrating the WSI-based risk score with clinicopathological parameters to construct multimodal recurrence risk prediction models. These models were developed using CPH, EN-Cox, GBRT, Lasso-Cox, Ridge-Cox, and RSF.

Given the limited number of HR+/HER2- EBC patients with complete clinical information ($n = 968$), we combined the training and validation sets from the deep learning five-fold cross-validation to form the machine learning training set, with the deep learning test set serving as the machine learning test set. The multimodal GBRT model exhibited superior predictive performance. In the test set, the AUC values for 3-, 5-, and 7-year predictions were 0.875 ± 0.037, 0.864 ± 0.031, and 0.866 ± 0.029, respectively. Other multimodal models also achieved robust performance, with AUC values consistently exceeding 0.8 across all time points (Fig. 6A). Due to its consistently high predictive performance, the



Multivariate Cox regression analysis

Characteristics	Num/Mean±SD(Recurrence)	HR (95%CI)	P.value
LVI	968		
-	886 (108)		
+	82 (22)	1.83 [1.13, 2.96]	0.014*
HER2			
0	263 (35)		
low	705 (95)	1.46 [0.98, 2.17]	0.064
Molecular subtype			
Luminal A	209 (25)		
Luminal B	759 (105)	0.83 [0.48, 1.42]	0.495
Tumor size			
≤ 2cm	414 (40)		
＞2cm	554 (90)	1.21 [0.82, 1.78]	0.344
LN status			
0	410 (40)		
1~3	364 (45)	1.30 [0.84, 2.02]	0.245
≥ 4	194 (45)	1.98 [1.27, 3.08]	0.003*
Grade			
1	39 (3)		
2	554 (53)	1.08 [0.33, 3.49]	0.904
3	375 (74)	1.39 [0.43, 4.54]	0.583
Age	48.58±9.29 (130)	1.01 [0.98, 1.03]	0.581
ER	0.82±0.17 (130)	0.41 [0.17, 1.02]	0.055
PR	0.60±0.32 (130)	0.62 [0.34, 1.13]	0.122
Ki67	0.26±0.17 (130)	1.12 [0.36, 3.48]	0.841
WSI score	0.11± 0.30(130)	10.35 [7.07, 15.16]	<0.001*

Num: Number; Recurrence: Number of recurrence.

Multivariate Cox regression analysis_fold 0

Characteristics	Num/Mean±SD(Recurrence)	HR (95%CI)	P.value
LVI	968		
-	886 (108)		
+	82 (22)	2.06 [1.28, 3.32]	0.003*
HER2			
0	263 (35)		
low	705 (95)	1.53 [1.03, 2.27]	0.036*
Molecular subtype			
Luminal A	209 (25)		
Luminal B	759 (105)	0.97 [0.57, 1.66]	0.915
Tumor size			
≤ 2cm	414 (40)		
＞2cm	554 (90)	1.25 [0.85, 1.83]	0.264
LN status			
0	410 (40)		
1~3	364 (45)	1.25 [0.81, 1.94]	0.312
≥ 4	194 (45)	2.02 [1.30, 3.13]	0.002*
Grade			
1	39 (3)		
2	554 (53)	1.13 [0.35, 3.64]	0.839
3	375 (73)	1.49 [0.46, 4.83]	0.502
Age	48.58±9.29 (130)	0.99 [0.97, 1.01]	0.395
ER	0.82±0.17 (130)	0.55 [0.22, 1.38]	0.200
PR	0.60±0.32 (130)	0.73 [0.41, 1.30]	0.287
Ki67	0.26±0.17 (130)	1.21 [0.38, 3.87]	0.742
WSI score			
Low risk	453 (8)		
High risk	515 (122)	12.67 [6.16, 26.09]	<0.001*

Fig. 5 Forest plot of multivariate Cox regression analysis. (**A**) Multivariate Cox regression analysis of WSI-based risk score as a continuous variable and clinicalpathological features for RFS. (**B**) Multivariate Cox regression analysis of WSI-based risk score as a categorical variable and clinicalpathological features for RFS in fold 0. WSI, whole slide image; RFS, recurrence-free survival. LN, lymph node; LVI: lymphatic vessel infiltration; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor

GBRT model was selected as the final prediction model. Further analysis revealed that the predictive performance of the WSI-based risk score alone surpassed that of clinicopathological features alone. In the test set, the AUC values for the WSI-based risk score were 0.840 ± 0.033 (3-year), 0.814 ± 0.078 (5-year), and 0.826 ± 0.066 (7-year), compared to 0.684 ± 0.064 (3-year), 0.650 ± 0.062 (5-year), and 0.641 ± 0.048 (7-year) for clinicopathological features (Fig. 6B). Moreover, the multimodal prediction model that combined the WSI-based risk score with clinicopathological features achieved average AUC improvements of 19.1%, 21.4%, and 22.5% for 3-, 5-, and 7-year predictions, respectively, compared to models relying solely on clinicopathological features (Fig. 6B).

To elucidate how the multimodal recurrence risk prediction model predicts recurrence in HR+/HER2- EBC patients, we utilized SHapley Additive exPlanations (SHAP) values to clarify the impact of each feature variable on the prediction models. Feature importance rankings from the SHAP summary plots of six different multimodal recurrence risk prediction models revealed that the WSI-based risk score was the most significant contributor across all models (Fig. 6C-H; additional results for another four folds were shown in Figure S11-S14). Among the other top five contributing variables were lymph node (LN) status, progesterone receptor (PR) status, histological grade, and HER2 status. Additionally, SHAP dependency analysis was used to explore how individual feature variables influence the outcomes of different predictive models (Figure S10-S14). A higher SHAP value for a feature variable indicates a greater likelihood of recurrence in HR+/HER2- EBC patients following adjuvant C-ET. For instance, in different models, a lower value of the WSI-based risk score corresponded to a negative SHAP value, which was associated with a reduced recurrence risk Conversely, a higher value of the WSI-based risk score corresponded to a positive SHAP value, indicating a stronger influence on the prediction of recurrence risk in HR+/HER2- EBC patients.

To enhance clinical utility, we developed a recurrence prediction tool for HR+/HER2- EBC patients undergoing adjuvant C-ET in the form of a nomogram (Fig. 7,

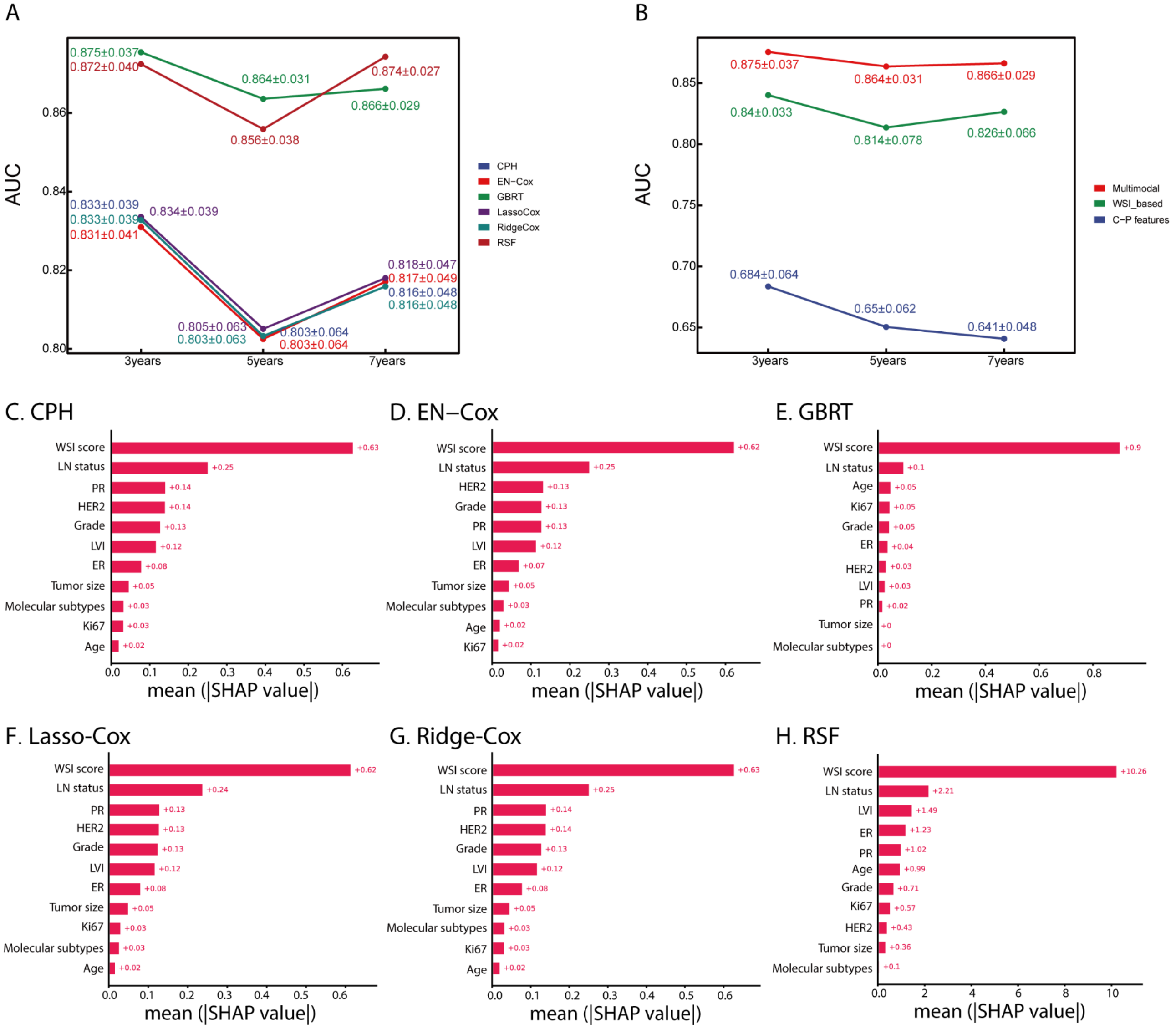


Fig. 6 Integrating WSI-based risk score and clinicalpathological factors to construct a multimodal recurrence risk prediction model. (**A-B**). Comparison of AUC among different multimodal recurrence risk prediction models. (**C-H**). SHAP summary plots for machine learning models in fold 0-4. Importance matrices and SHAP summary plots illustrated the contribution of feature variables to different models.WSI, whole slide image; LN, lymph node; LVI: lymphatic vessel infiltration; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor

Figure S15). The nomogram was constructed using the top five feature variables identified by feature importance ranking across the six models (WSI-based risk score, LN status, PR, histological grade and HER2 status). Each feature variable in the nomogram is associated with a score, and the total score, calculated by summing the individual scores for all feature variables, represents the predicted recurrence risk for HR+/HER2- EBC patients.

Discussion

In recent years, there has been a trend towards emphasizing de-escalated treatments in adjuvant therapy for HR+/HER2- EBC, supported by studies demonstrating the feasibility of exempting chemotherapy for certain HR+/HER2- breast cancer patients [16, 17]. However, approximately 20% of HR+/HER2- EBC patients still experience recurrence despite undergoing adjuvant C-ET [2]. With the advent of new treatment regimens, accurately predicting the recurrence risk in these patients is essential for optimizing therapeutic strategies and identifying candidates who may benefit from more tailored treatment approaches. In this study, we developed two deep learning pipelines (ACMIL-based and CLAM-based) that integrate multi-instance learning frameworks with tissue-specific feature encoders to accurately predict recurrence risk in HR+/HER2- EBC patients undergoing adjuvant C-ET. Our results demonstrate that the

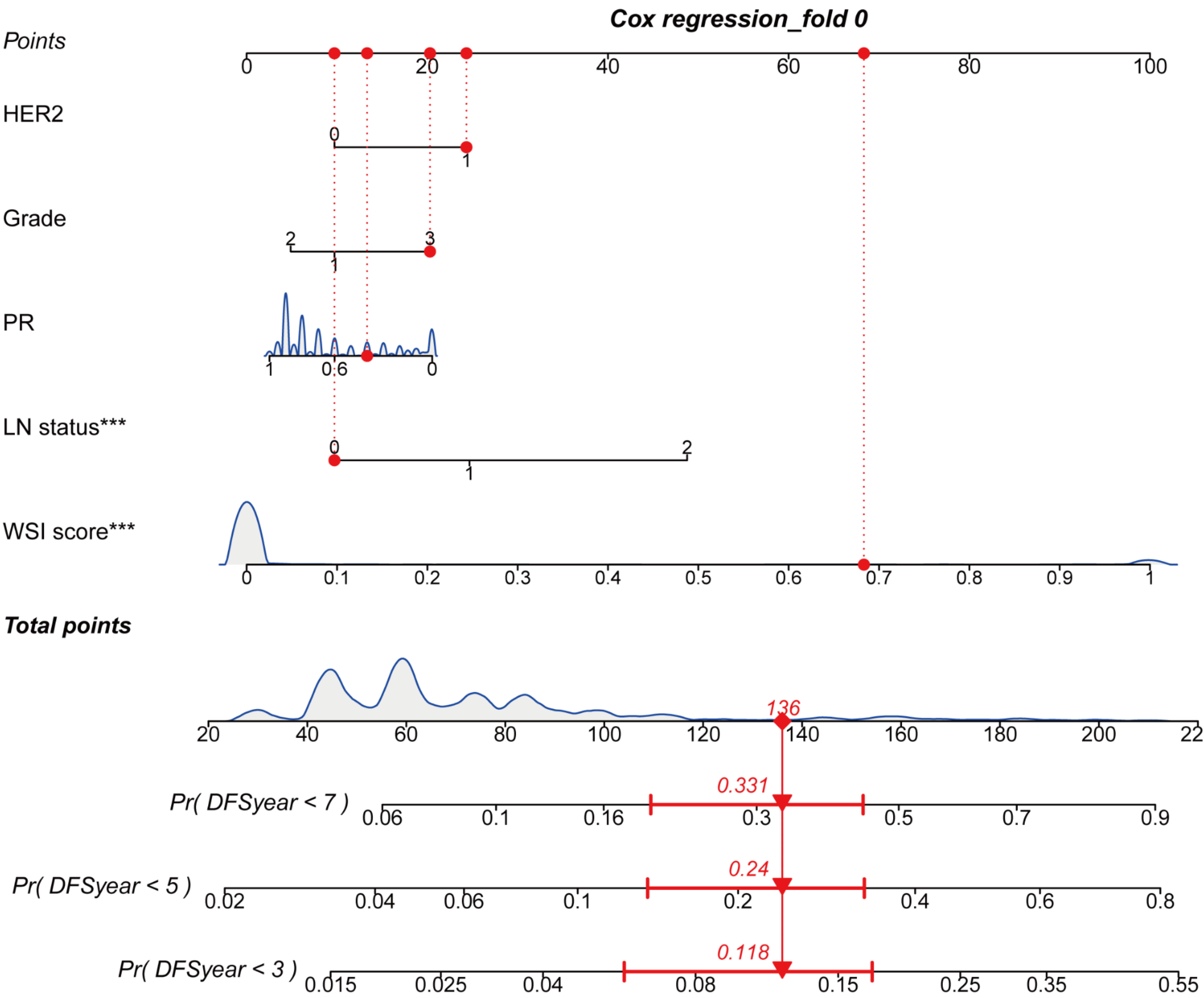


Fig. 7 Representative nomogram for predicting recurrence risk in HR+/HER2- EBC patients undergoing adjuvant C-ET based on WSI-based risk score, LN status, LIV status, PR and HER2 in fold 0. EBC, early breast cancer; C-ET, chemo-endocrine therapy; WSI, whole slide image, LN, lymph node; LVI: lymphatic vessel infiltration; PR, progesterone receptor; HER2, human epidermal growth factor receptor

proposed multimodal prediction model offers significant advantages in both technical performance and clinical utility.

First, regarding the selection of feature encoders, we systematically evaluated the performance of five pre-trained feature encoders (CTransPath [9], CONCH [10], UNI [12], REMEDIS [11], and Virchow [13]). Although all these models achieved high predictive performance, the UNI encoder consistently outperformed the others across multiple metrics, including AUC, accuracy, recall, and F1 score, particularly within the ACMIL pipeline. This finding underscores the critical role of feature encoder selection in enabling the model to effectively capture tissue-specific features and manage the complexity of pathological images. The DINOv2-based self-supervised learning method employed by UNI demonstrated unique advantages in feature diversity, suggesting that future improvements in feature encoders could benefit from advanced training strategies such as contrastive learning or other self-supervised approaches. Furthermore, the compatibility between feature encoders and downstream models emerged as a crucial factor. Selecting appropriate feature encoders is essential for optimizing the overall performance of the model.

In the evaluation of deep learning pipeline performance, the UNI-ACMIL pipeline demonstrated exceptional predictive capability, achieving AUCs of 0.86 ± 0.02 in five-fold cross-validation test sets and 0.80 ± 0.04 in external validation. Existing WSI-based prognostic models for HR+/HER2- breast cancer primarily focus on recurrence risk prediction using OncotypeDX scores [18–20]. For example, Howard et al. reported an AUC of 0.798 for deep learning-based pathological models and 0.828 for models integrating clinical and pathological features [18]. In addition, the recently published TITAN model from the Harvard group achieved a C-index of

0.757 ± 0.015 for survival prediction across all breast cancer subtypes [21]. In comparison, our model achieved a C-index of 0.837 ± 0.023 for HR+/HER2- EBC in the WCH cohort, demonstrating superior performance. Moreover, incorporating manually annotated ROIs further enhanced model outcomes, indicating that precise tumor region annotation plays a vital role in extracting critical features and improving predictive accuracy. However, manual annotation is labor-intensive and subject to variability, underscoring the need for automated annotation tools to reduce human intervention, improve consistency, and facilitate large-scale multicenter studies.

From a clinical perspective, the multimodal prediction model in this study, which combines WSI-based risk scores with clinicopathological features, provided accurate recurrence risk assessment. The high prognostic value of the multimodal model was demonstrated in the five-fold cross-validation test sets, with AUCs of 0.875 ± 0.037, 0.864 ± 0.031, and 0.866 ± 0.029 for recurrence prediction at 3-, 5-, and 7-years, respectively. Compared to models relying solely on clinicopathological features, the multimodal model improved prediction performance by 19.1-22.5%. Additionally, SHAP analysis consistently identified the WSI-based risk score as the most significant variable across all models, further emphasizing its clinical relevance.

This study identified the pathological and molecular characteristics of high-risk patients through an integrated analysis of morphological features from pathological slides and patient transcriptomic data. WSIs from high-risk patients exhibited cord-like or sheet-like distributions of tumor cells with prominent nucleoli, enlarged nuclei, and chromatin condensation. Immune microenvironment analysis revealed lower proportions of gamma-delta T cells and higher proportions of monocytes in high-risk patients, indicating that the immune microenvironment plays a crucial role in tumor recurrence [22–24]. GSEA revealed that differentially expressed genes in high-risk patients were significantly enriched in pathways such as IFN-α/γ response, transplant rejection, and KRAS signaling, which may contribute to recurrence mechanisms.

Despite the promising potential demonstrated by this study, several limitations must be addressed. First, the patient cohort primarily consisted of single-center data. Although external validation results were favorable, multicenter studies with larger sample sizes are necessary to assess the model's generalizability. Second, while the WSI-based risk score was the most significant contributor to the multimodal model, its underlying biological mechanisms and interactions with other clinical features require further exploration. Future research should focus on integrating advanced deep learning techniques, such as UNet or DeepLab, to enable automated, high-precision tumor region annotation. Additionally, investigating the model's potential applications in clinical decision-making, such as guiding the use of CDK4/6 inhibitors [25–27] or other targeted therapies, would further enhance its utility.

In conclusion, the proposed deep learning-based recurrence risk prediction model demonstrates significant clinical potential for HR+/HER2- EBC patients. By integrating multimodal data, the model substantially improves predictive performance, offering a novel tool for personalized treatment and precision medicine. Further optimization of feature encoders and interpretability will promote the broader adoption of deep learning technologies in pathological image analysis and advance the field of precision oncology.

Abbreviations

EBC	Eearly breast cancer
C-ET	Chemo-endocrine therapy
AUC	The area under the curve
ET	Endocrine therapy
WCH	West china hospital
RFS	Recurrence-free survival
OS	Overall survival
CI	Confidence interval
ROC	Receiver operating characteristic
DEGs	Differentially expressed genes
GSEA	Gene set enrichment analysis
LN	Lymph node
LVI	Lymphatic vessel invasion
PR	Progesterone receptor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13058-025-01968-0 .

Supplementary Material 1

Acknowledgements

The author thanks the West China Biobanks, West China Hospital, Sichuan University, for providing fresh frozen breast cancer tissues for transcriptome sequencing.

Author contributions

XY, JL, JC, HB and YH contributed to the study design and conception. XY and WC contributed to the data selection. XY, YM, JL, JC, MZ and YH contributed to the model development and predictions. XY, YM and XX contributed to the data analysis. XY wrote the first draft of the manuscript. XY and YH. reviewed and edited the manuscript. HB and YH contributed to funding acquisition. All authors read and approved the final manuscript.

Funding

The study was supported by the Fundamental Research Funds for the Central Universities, under grant Sichuan University (YJ202285) and the 1·3·5 project for disciplines of excellence (ZYGD18012).

Data availability

The data that support the findings of this study are available on request from the corresponding author.

Declarations

Ethics approval and consent to participate

Our study was approved by the ethical committee of West China Hospital, Sichuan University (No.2365 in 2023), and abided with the Declaration of Helsinki before using tissue samples for scientific researches purpose only. The written informed consent was waived by the ethical committee for this retrospective study.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Pathology, West China Hospital, Sichuan University, Chengdu, China
[2]Institute of Clinical Pathology, West China Hospital, Sichuan University, Chengdu 610041, China
[3]College of Computer Science, Sichuan University, Chendu, China
[4]Institute for Breast Health Medicine, Cancer Center, Breast Center, West China Hospital, Sichuan University, Chengdu, China

Received: 30 July 2024 / Accepted: 21 January 2025
Published online: 28 March 2025

References

1. Slamon DJ, Fasching PA, Hurvitz S, Chia S, Crown J, Martin M, Barrios CH, Bardia A, Im SA, Yardley DA, et al. Rationale and trial design of NATALEE: a phase III trial of adjuvant ribociclib + endocrine therapy versus endocrine therapy alone in patients with HR+/HER2- early breast cancer. Ther Adv Med Oncol. 2023;15:17588359231178125.
2. Early Breast Cancer Trialists', Collaborative G, Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, McGale P, Pan HC, Taylor C, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet. 2011;378(9793):771–84.
3. Mercan C, Balkenhol M, Salgado R, Sherman M, Vielh P, Vreuls W, Polonia A, Horlings HM, Weichert W, Carter JM, et al. Deep learning for fully-automated nuclear pleomorphism scoring in breast cancer. NPJ Breast Cancer. 2022;8(1):120.
4. Saha M, Chakraborty C, Arun I, Ahmed R, Chatterjee S. An Advanced Deep Learning Approach for Ki-67 stained Hotspot Detection and Proliferation Rate Scoring for Prognostic evaluation of breast Cancer. Sci Rep. 2017;7(1):3213.
5. Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, Rohr K, Shah MA, Wang D, Rousson M, et al. Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge. Med Image Anal. 2019;54:111–21.
6. Jiang X, Hoffmeister M, Brenner H, Muti HS, Yuan T, Foersch S, West NP, Brobeil A, Jonnagaddala J, Hawkins N, et al. End-to-end prognostication in colorectal cancer by deep learning: a retrospective, multicentre study. Lancet Digit Health. 2024;6(1):e33–43.
7. Gui CP, Chen YH, Zhao HW, Cao JZ, Liu TJ, Xiong SW, Yu YF, Liao B, Cao Y, Li JY, et al. Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study. Lancet Digit Health. 2023;5(8):e515–24.
8. Fremond S, Andani S, Barkey Wolf J, Dijkstra J, Melsbach S, Jobsen JJ, Brinkhuis M, Roothaan S, Jurgenliemk-Schulz I, Lutgens L, et al. Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts. Lancet Digit Health. 2023;5(2):e71–82.
9. Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, Huang J, Han X. Transformer-based unsupervised contrastive learning for histopathological image classification. Med Image Anal. 2022;81:102559.
10. Lu MY, Chen B, Williamson DFK, Chen RJ, Liang I, Ding T, Jaume G, Odintsov I, Le LP, Gerber G, et al. A visual-language foundation model for computational pathology. Nat Med. 2024;30(3):863–74.
11. Azizi S, Culp L, Freyberg J, Mustafa B, Baur S, Kornblith S, Chen T, Tomasev N, Mitrovic J, Strachan P, et al. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. Nat Biomed Eng. 2023;7(6):756–79.
12. Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, Chen B, Zhang A, Shao D, Shaban M, et al. Towards a general-purpose foundation model for computational pathology. Nat Med. 2024;30(3):850–62.
13. Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Severson K, Zimmermann E, Hall J, Tenenholtz N, Fusi N, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nat Med. 2024;30(10):2924–35.
14. Zhang Y, Li H, Sun Y, Zheng S, Zhu C, Yang L. Attention-Challenging Multiple Instance Learning for Whole Slide Image Classification. *ArXiv* 2023, abs/2311.07125.
15. 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(6):555–70.
16. Piccart M, van 't Veer LJ, Poncet C, Lopes Cardozo JMN, Delaloge S, Pierga JY, Vuylsteke P, Brain E, Vrijaldenhoven S, Neijenhuis PA, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. Lancet Oncol. 2021;22(4):476–88.
17. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, Lin NU, Perez EA, Goldstein LJ, Chia SKL, et al. 21-Gene assay to inform Chemotherapy Benefit in Node-positive breast Cancer. N Engl J Med. 2021;385(25):2336–47.
18. Howard FM, Dolezal J, Kochanny S, Khramtsova G, Vickery J, Srisuwananukorn A, Woodard A, Chen N, Nanda R, Perou CM, et al. Integration of clinical features and deep learning on pathology for the prediction of breast cancer recurrence assays and risk of recurrence. NPJ Breast Cancer. 2023;9(1):25.
19. Li H, Wang J, Li Z, Dababneh M, Wang F, Zhao P, Smith GH, Teodoro G, Li M, Kong J, et al. Deep learning-based Pathology Image Analysis enhances Magee feature correlation with Oncotype DX breast recurrence score. Front Med (Lausanne). 2022;9:886763.
20. Cho SY, Lee JH, Ryu JM, Lee JE, Cho EY, Ahn CH, Paeng K, Yoo I, Ock CY, Song SY. Author correction: deep learning from HE slides predicts the clinical benefit from adjuvant chemotherapy in hormone receptor-positive breast cancer patients. Sci Rep. 2021;11(1):21043.
21. Ding T, Wagner SJ, Song AH, Chen RJ et al. Multimodal Whole Slide Foundation Model for Pathology, Arxiv, 2024.
22. Valenza C, Taurelli Salimbeni B, Santoro C, Trapani D, Antonarelli G, Curigliano G. Tumor infiltrating lymphocytes across breast Cancer subtypes: current issues for Biomarker Assessment. Cancers (Basel) 2023; 15(3).
23. Mamedov MR, Vedova S, Freimer JW, Sahu AD, Ramesh A, Arce MM, Meringa AD, Ota M, Chen PA, Hanspers K, et al. CRISPR screens decode cancer cell pathways that trigger gammadelta T cell detection. Nature. 2023;621(7977):188–95.
24. Hu Y, Hu Q, Li Y, Lu L, Xiang Z, Yin Z, Kabelitz D, Wu Y. Gammadelta T cells: origin and fate, subsets, diseases and immunotherapy. Signal Transduct Target Ther. 2023;8(1):434.
25. Slamon D, Lipatov O, Nowecki Z, McAndrew N, Kukielka-Budny B, Stroyakovskiy D, Yardley DA, Huang CS, Fasching PA, Crown J, et al. Ribociclib plus Endocrine Therapy in early breast Cancer. N Engl J Med. 2024;390(12):1080–91.
26. Rastogi P, O'Shaughnessy J, Martin M, Boyle F, Cortes J, Rugo HS, Goetz MP, Hamilton EP, Huang CS, Senkus E, et al. Adjuvant Abemaciclib Plus endocrine therapy for hormone Receptor-Positive, human epidermal growth factor receptor 2-Negative, high-risk early breast Cancer: results from a preplanned monarchE overall survival interim analysis, including 5-Year efficacy outcomes. J Clin Oncol. 2024;42(9):987–93.
27. Johnston SRD, Toi M, O'Shaughnessy J, Rastogi P, Campone M, Neven P, Huang CS, Huober J, Jaliffe GG, Cicin I, et al. Abemaciclib plus endocrine therapy for hormone receptor-positive, HER2-negative, node-positive, high-risk early breast cancer (monarchE): results from a preplanned interim analysis of a randomised, open-label, phase 3 trial. Lancet Oncol. 2023;24(1):77–90.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.