Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Habitat-Based Multimodal Deep Learning for Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer: A Two-Center Multiparametric MRI Study

Xiancong Gao
Suzhou Hospital Affiliated to Nanjing Medical University
Xun Wang
Suzhou Hospital Affiliated to Nanjing Medical University
Pei Cai
Affiliated Suqian Hospital of Xuzhou Medical University
Jiahai Huang
Suzhou Hospital Affiliated to Nanjing Medical University
Quanquan Guo
Affiliated Suqian Hospital of Xuzhou Medical University
Liyou Huang
huangliyoua@outlook.com

Affiliated Suqian Hospital of Xuzhou Medical University

Research Article

Keywords: Breast cancer, axillary lymph node metastasis, multiparametric MRI, habitat radiomics, deep learning

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11496762/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
No competing interests reported.

Table 1 is available in the Supplementary Files section.

Habitat-Based Multimodal Deep Learning for Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer: A Two-Center Multiparametric MRI Study

Xiancong Gao[a,b,1], Xun Wang[c,1], Pei Cai[a,b], Jiahai Huang[c], Quanquan Guo[d,e], Liyou Huang[d,e,*]

1 These authors contributed equally to this work.

a Department of Radiology, Affiliated Suqian Hospital of Xuzhou Medical University, Suqian 223800, PR China

b Department of Radiology, Nanjing Drum Tower Hospital Group Suqian Hospital, Suqian 223800, PR China

c Department of Radiology, Suzhou Hospital Affiliated to Nanjing Medical University, Suzhou 215000, PR China

d Department of Oncology, Affiliated Suqian Hospital of Xuzhou Medical University, Suqian 223800, PR China

e Department of Oncology, Nanjing Drum Tower Hospital Group Suqian Hospital, Suqian 223800, PR China

* Corresponding author: Liyou Huang, Department of Oncology, Affiliated Suqian Hospital of Xuzhou Medical University, Suqian 223800, PR China. Tel: +86 13141561776; E-mail: huangliyoua@outlook.com

Background: Axillary lymph node (ALN) status is a critical determinant of breast cancer staging and treatment planning. This study aimed to develop and validate a habitat-based

multimodal deep learning (HMDL) model that learns whole-tumor and habitat subregion-level multiparametric MRI features, radiomics features, and clinical features within a unified network to enable noninvasive preoperative prediction of ALN metastasis in breast cancer.

Methods: This retrospective study enrolled 201 patients with breast cancer from two centers. At Center 1, 160 patients were randomly divided into training (n = 128) and internal validation (n = 32) sets at the patient level in an 8:2 ratio. The remaining 41 patients from Center 2 served as an independent external validation set. A two-stage 3D U-Net was used for automatic segmentation of the breast tumor region of interest (ROI). Based on dynamic contrast-enhanced MRI (DCE), fat-suppressed T2-weighted imaging (T2FS), and diffusion-weighted imaging (DWI), K-means clustering was applied to partition intratumoral voxels into habitat subregions, and radiomics features were extracted from the whole tumor and each subregion. The HMDL model used a 2D convolutional neural network (CNN) as the backbone, extracting whole-tumor and habitat subregion image features along three orthogonal planes, fusing them via cross-modal and cross-region multi-head self-attention into a unified imaging representation, and concatenating it with radiomics and clinical features for classification. Ablation experiments were performed to evaluate the contribution of each information source, input sequence, and input strategy. Clinical and clinical-habitat radiomics models were constructed as comparators, and DeLong tests were used to compare the area under the receiver operating characteristic curve (AUC) among the three models.

Results: The automatic segmentation model achieved Dice similarity coefficients of 0.842 and 0.804 on the internal and external validation sets, respectively. Based on the Calins

ki-Harabasz index, each tumor was partitioned into 3 habitat subregions. Ablation experiments showed that removing habitat subregion representations, radiomics features, clinical features, or any single sequence decreased the AUC on both validation sets. Among the three sequences, removal of DWI caused the largest AUC decrease in the HMDL (w/o DWI) model (AUC, 0.702 and 0.676, respectively), with statistically significant differences (P = 0.015, P = 0.012). The HMDL model achieved AUCs of 0.929 and 0.913 on the internal and external validation sets, respectively, both higher than the clinical-habitat radiomics model (0.869 and 0.841) and the clinical model (0.736 and 0.709). The difference between the HMDL and clinical models was statistically significant on the external validation set (P = 0.027).

Conclusions: The HMDL model enables noninvasive quantitative prediction of ALN metastasis in breast cancer by jointly optimizing whole-tumor and habitat subregion-level imaging information, radiomics features, and clinical features within a unified network end-to-end, and demonstrates robust performance in independent external validation, offering a novel technical approach for preoperative ALN assessment.

Keywords: Breast cancer; axillary lymph node metastasis; multiparametric MRI; habitat radiomics; deep learning

Background

Breast cancer is the most prevalent malignancy among women worldwide [1]. Axillary lymph node (ALN) metastasis status is a pivotal factor in determining breast cancer staging, treatment strategy, and prognosis. The 10-year follow-up results of the American College of Surgeons Oncology Group (ACOSOG) Z0011 randomized clinical trial demonstrate

d that axillary lymph node dissection (ALND) did not confer an overall survival benefit i n patients with breast-conserving surgery and one to two sentinel lymph node macrometa stases but significantly increased the risk of complications such as upper extremity lymph edema [2]. Therefore, accurate noninvasive preoperative prediction of ALN metastasis is essential to avoid overtreatment. Multiparametric magnetic resonance imaging (mpMRI) has become a standard tool for preoperative local staging of breast cancer; however, the d iagnostic efficacy of conventional visual assessment for ALN status remains limited and i s insufficient to guide individualized axillary management.

Radiomics, by extracting high-throughput quantitative imaging features, provides an effe ctive tool for noninvasive characterization of tumor phenotypes [3]. Building upon this, h abitat imaging quantifies intratumoral spatial heterogeneity by partitioning the tumor into functionally distinct subregions. Studies have shown that habitat subregion information p rovides incremental predictive value for ALN metastasis [4,5]. However, these studies rel ied on predefined formulaic handcrafted features, whose expressive dimensions are const rained by prior assumptions of low-order statistics and texture descriptors, making it diffi cult to fully capture the complex high-throughput imaging patterns associated with lymph node metastasis. Deep learning overcomes the limitations of handcrafted feature design b y automatically learning hierarchical imaging representations and has demonstrated subst antial potential in predicting breast ALN metastasis [6]. The complementary information provided by mpMRI can further enhance the predictive performance of deep learning mo dels for breast ALN metastasis [7]. Nevertheless, most existing deep learning studies treat the tumor as a homogeneous whole input to the network without characterizing signal dif

ferences at the subregional level, thereby failing to reflect the spatial heterogeneity associated with ALN metastasis.

To address this gap, we propose a habitat-based multimodal deep learning (HMDL) model. This study aims to partition tumors into functionally consistent habitat subregions through multimodal voxel clustering, and subsequently construct a convolutional neural network (CNN) architecture incorporating intra-modal region fusion and cross-modal, cross-region multi-head self-attention mechanisms. We anticipate that this deep learning model will collaboratively learn habitat subregion representations and multisequence imaging features to enable accurate noninvasive preoperative prediction of breast cancer ALN metastasis, providing more effective support for individualized treatment decision-making.

Methods

Patients

This retrospective study collected data from patients with breast cancer who underwent preoperative multiparametric breast MRI at two hospitals (Figure 1). The inclusion criteria were as follows: [1] female patients with histopathologically confirmed primary invasive breast cancer; [2] completion of multiparametric breast MRI within 2 weeks before surgery, including full DCE, T2FS, and DWI sequences; [3] patients who underwent sentinel lymph node biopsy (SLNB) or axillary lymph node dissection (ALND) with definitive ALN pathologic diagnosis. The exclusion criteria were as follows: [1] bilateral breast cancer, multifocal lesions, or distant metastasis at initial diagnosis; [2] history of neoadjuvant chemotherapy, radiotherapy, endocrine therapy, or prior breast surgery before MRI; [3] images with severe motion artifacts, metal artifacts, or missing sequences affecting image qu

ality; [4] indistinct tumor boundaries precluding accurate delineation of the region of inte rest (ROI).

A total of 201 patients met the inclusion criteria. Data from Center 1 (Suzhou Municipal Hospital) included 160 patients collected between January 2020 and October 2025, rando mly divided into a training set (n = 128) and an internal validation set (n = 32) in an 8:2 r atio at the patient level. Data from Center 2 (Affiliated Suqian Hospital of Xuzhou Medic al University) included 41 patients collected between January 2022 and January 2026, ser ving as an independent external validation set. The overall technical workflow is illustrat ed in Figure 2.

This study was conducted in accordance with the Declaration of Helsinki and was approv ed by the Ethics Committee of Affiliated Suqian Hospital of Xuzhou Medical University (approval number: 2023038). Because of its retrospective design, the requirement for info rmed consent was waived.

MRI Image Acquisition

MRI images were retrieved from the hospital picture archiving and communication syste m (PACS), and clinical and pathologic parameters were extracted from the electronic me dical record system, including patient age, menopausal status, axillary lymph node (ALN) status, estrogen receptor (ER), progesterone receptor (PR), human epidermal growth fact or receptor 2 (HER2), and Ki-67 index. According to pathology reports, ER, PR, HER2, a nd Ki-67 were assessed on preoperative core needle biopsy specimens; ER and PR positi vity was defined as immunohistochemical staining in ≥ 10% of tumor cells; HER2 status was determined by immunohistochemistry (IHC), with equivocal cases (2+) further evalu

ated by fluorescence in situ hybridization (FISH); Ki-67 index ≥ 14% was considered positive. ALN status was determined by the surgical pathologic results of SLNB or ALND specimens. ALN metastasis positivity was defined as the presence of macrometastases (> 2 mm) or micrometastases (0.2–2 mm). Patients were classified into ALN metastasis-positive and ALN metastasis-negative groups.

At Center 1, patients were scanned on two scanners: a Siemens 1.5T MRI system (Aera) and a Philips 3.0T MRI system (Ingenia). At Center 2, all patients were scanned on a Philips 3.0T MRI system (Ingenia). A dedicated bilateral breast phased-array coil was used. Patients were positioned prone with both breasts suspended naturally within the coil. Scanning protocols for both 1.5T and 3.0T MRI included T1-weighted imaging (T1WI), fat-suppressed T2-weighted imaging (T2FS), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced MRI (DCE). Detailed scan parameters are provided in Additional file 1.

Tumor Segmentation

In this study, the second post-contrast phase of DCE was used [8], and the highest b-value image was selected for DWI. N4 bias field correction was independently applied to DCE, T2FS, and DWI sequences to eliminate low-frequency signal drift caused by radiofrequency field inhomogeneity. Z-score intensity standardization was performed to reduce intensity scale differences across scanners. All sequences were resampled to isotropic 1 × 1 × 1 mm^3 voxels using trilinear interpolation. Training set DCE images were imported into ITK-SNAP (version 3.8.0; http://www.itksnap.org). A radiologist with 7 years of breast MRI diagnostic experience delineated the breast region and intratumoral ROI on DCE im

ages slice by slice, ensuring inclusion of all enhancing regions and exclusion of nonenhancing necrotic areas and adjacent vascular structures. A second radiologist with 10 years of breast MRI diagnostic experience reviewed and finalized the delineations. Subsequent review and correction of all automatic segmentation results and visual inspection of registration results were performed by these two radiologists. To augment the segmentation model training data, an additional 4 patients from Center 1 and 15 patients from Center 2 whose T2FS or DWI sequences were missing and therefore not included in downstream analysis were enrolled; their DCE images underwent the same preprocessing and ROI delineation workflow and were included in the two-stage segmentation model training but were not used for subsequent radiomics feature extraction or classification modeling.

A two-stage segmentation strategy was adopted: first, localizing the breast region to exclude background interference from the chest wall and heart; second, performing fine segmentation of the intratumoral ROI within the constrained search space. Both stages used a 3D U-Net multi-scale attention variant as the backbone network; detailed model descriptions are provided in Additional file 2. The two stages shared the same backbone architecture but were trained independently; the output of the first stage, after expert correction, served as the input constraint for the second stage. All segmentation results were reviewed and corrected slice by slice by the aforementioned two radiologists.

The first stage was breast region segmentation, aiming to segment breast tissue from the entire MRI image and exclude irrelevant anatomical structures such as the chest wall, lungs, and heart. The network took the 3D volume of the DCE sequence as input, and the output breast binary mask served as the search-space constraint for the second stage. The second stage was intratumoral ROI segmentation, aiming to accurately segment the tumor wit

hin the breast region. A cascaded training strategy was employed: first, a low-resolution model was trained on images resampled to 1.5-mm voxel spacing, and then its predicted probability map was concatenated with the original DCE image along the channel dimens ion and fed into the original-resolution high-resolution cascaded model for fine segmentat ion. The network input was the DCE image with background masked by the first-stage br east mask (voxels outside the mask were set to zero). Segmentation accuracy was assesse d using the Dice similarity coefficient by comparing automatically segmented ROIs with radiologist-adjusted ROIs on the internal and external validation sets. Hyperparameters fo r both stages are provided in Additional file 3.

Multisequence Registration and Habitat Partitioning

Multisequence registration used DCE as the fixed reference, registering T2FS and DWI t o the DCE space. Because the three MRI sequences of each patient were acquired during the same examination, they were already partially aligned. Before registration, the spatial sampling grids of T2FS and DWI were aligned with DCE, and the normalized mutual inf ormation (NMI) of this alignment relative to DCE was computed as the baseline NMI. A two-stage progressive registration strategy from global to local was then adopted; after ea ch stage, whether to retain the stage result was determined by NMI increment: if no impr ovement was observed, the process reverted to the previous stage, ensuring monotonic im provement. Stage 1 was global rigid registration using whole-image global NMI as the si milarity metric. Stage 2 was local affine fine-tuning based on the tumor ROI mask, using local NMI within the ROI as the similarity metric, with emphasis on optimizing tumor re gion alignment accuracy. After registration, all results were visually inspected slice by sli ce by the two aforementioned radiologists using ITK-SNAP, and manual adjustment was

performed via ITK-SNAP manual registration when necessary. K-means clustering was u sed to partition intratumoral voxels into functional subregions. To determine the optimal cluster number k, the Calinski-Harabasz (CH) index was computed across the range of 2 t o 5 [9]. After clustering, cluster labels were assigned to each voxel under the tumor ROI mask constraint, generating binary masks for k subregions.

PyRadiomics was used to extract radiomics features from the whole tumor and each of th e k habitat subregions across all three sequences. Each region yielded an 851-dimensional feature vector (including 162 first-order features, 14 shape features, 216 gray-level co-oc currence matrix features, 144 gray-level run-length matrix features, 144 gray-level size z one matrix features, 45 gray-level neighborhood difference matrix features, and 126 gray-level dependence matrix features). Because cross-patient cluster labels from unsupervised clustering lack biologically meaningful correspondence, the mean of each patient's subre gion features was taken as the patient-level habitat radiomics feature [10].

Habitat-Based Multimodal Deep Learning Model Development

The HMDL model developed in this study integrates multimodal information and regiona l heterogeneity. It uses a 2D CNN as the encoder; DCE, T2FS, and DWI each use indepe ndent encoders to preserve modality-specific feature expressions, with each modality pro cessing whole-tumor and k subregion images, and subregion images within the same mod ality sharing encoder weights. Three 2D backbone networks—ResNet-18, MobileNet-V2, and DenseNet-121—were compared to identify the optimal backbone.

Minimum bounding cube cropping was applied to each sequence image based on the who le-tumor mask. Pixel intensities outside the whole-tumor region were set to the minimum

intensity value of the image to eliminate extratumoral information. A 2D three-plane input strategy was adopted to fully exploit the 3D spatial information of the tumor: the slice with the largest cross-sectional area (axial) of the whole-tumor mask was used as the center, and the two adjacent slices were combined to form a three-channel 2D whole-tumor image. To compensate for the spatial information loss of a single-plane view, the same method was used to generate coronal and sagittal maximum-section images, so that images from three orthogonal planes were jointly input to the network, enabling the network to learn structural information from multiple directions. Based on the whole-tumor images, habitat subregion images for each slice were generated by element-wise multiplication with the corresponding subregion binary mask. All images were uniformly resized to 224 × 224 pixels for network input.

Encoder output features underwent intra-modal region fusion: subregion features were average-pooled and concatenated with whole-tumor features, and the concatenated feature vector was passed through a fully connected layer for dimensionality reduction to a unified single-modality feature. Subsequently, features from the three modalities entered a cross-modal, cross-region multi-head self-attention module for intermodal feature fusion, and the output was mean-pooled to obtain a unified fused feature. During classification, the fused feature was concatenated with radiomics features and clinical features, both processed by independent hierarchical multilayer perceptrons (MLPs), at the classifier entrance; after three-way fusion, the combined feature was fed into an MLP classifier to output binary logits. For the same case, the three planes were independently predicted, and the mean logits were passed through a sigmoid function to obtain the final patient-level metastasis probability.

In addition to the maximum section and its adjacent slices (max - 1, max, max + 1) from t hree directions, two additional sets of three-slice combinations centered at slices offset by ± 3 from the maximum section—(max + 2, max + 3, max + 4) and (max - 4, max - 3, ma x - 2)—were generated to augment the training sample size; augmented samples were use d only for training and were not included in prediction aggregation during evaluation. Tra nsfer learning with ImageNet pretrained weights, data augmentation, dropout regularizati on, weight decay, and early stopping were employed to prevent overfitting. HMDL mode l hyperparameters are provided in Additional file 4.

Ablation Experiments and Model Comparison

To validate the rationale of the 2D three-plane input strategy in the HMDL model, two co mparator models were constructed and compared with the HMDL model. First, a 2D sing le-plane comparator model HMDL (single-plane 2D) retaining only axial plane input was built, with identical network architecture, fusion modules, and training parameters. Seco nd, the 2D backbone encoder was replaced with the corresponding 3D version to construc t a 3D comparator model HMDL (3D), taking the minimum-bounding-cube cropped three -modality whole-tumor and subregion volumes (64 × 64 × 64) as input with randomly init ialized weights, performing end-to-end feature extraction and classification on 3D volum es; the fusion and classification modules remained consistent with the 2D framework.

Furthermore, to evaluate the necessity of each functional module in the HMDL model, ab lation experiments were conducted along three dimensions—imaging sequence, habitat re gion, and input information source (ablation models were named w/o X, where w/o denot es "without," indicating removal of the corresponding module X); all ablation configurati

ons were retrained from scratch using hyperparameters identical to the full model. At the imaging sequence level, the encoder of a single modality (DCE, T2FS, or DWI) was rem oved to construct HMDL (w/o DCE), HMDL (w/o T2FS), and HMDL (w/o DWI) model s. At the habitat region level, all subregion images and habitat radiomics features were re moved to construct HMDL (w/o habitat). Gradient-weighted class activation mapping (G rad-CAM) was used to visualize the model's attention regions to investigate whether diffe rent sequences and habitat subregion information were effectively utilized by the model.

At the information source level, radiomics features and clinical features were removed to construct HMDL (w/o radiomics) and HMDL (w/o clinical) models, respectively.

On the training set, the least absolute shrinkage and selection operator (LASSO) was used to reduce the dimensionality of clinical features; features with nonzero regression coeffic ients were retained via 5-fold cross-validation. Subsequently, three machine learning algo rithms—XGBoost, random forest (RF), and support vector machine (SVM)—were used t o construct models based on the selected features, and the algorithm with the highest area under the curve (AUC) on the external validation set was chosen as the final algorithm fo r the clinical model [11,12]. Hyperparameter search strategies for the three algorithms are provided in Additional file 5.

For radiomics features, clinical features, whole-tumor radiomics features, and habitat radi omics features were concatenated; features with a Pearson correlation coefficient > 0.9 w ere removed, and Z-score standardization was performed. The same feature selection and modeling workflow as the clinical model was then applied to construct the clinical-habita

t radiomics model. Differences in AUC among the HMDL model, clinical-habitat radiomics model, and clinical model were compared using the DeLong test.

Statistical Analysis

All statistical analyses were performed using R (version 4.3.0). Continuous variables were first assessed for normality using the Shapiro-Wilk test; those following a normal distribution were compared among multiple groups using one-way analysis of variance (ANOVA), and those not following a normal distribution were compared using the Kruskal-Wallis H test. Categorical variables were compared using the Pearson χ^2 test. Predictive model performance was evaluated using receiver operating characteristic (ROC) curves and AUC, accuracy, sensitivity, specificity, and F1 score. Differences in AUC between models were assessed using the DeLong test. A two-tailed P value < 0.05 was considered statistically significant.

Results

Patients' Clinical Characteristics

The clinical characteristics of the patients are summarized in Table 1. A total of 201 patients were enrolled, of whom 83 (41.3%) had ALN metastasis and 118 (58.7%) did not. No statistically significant differences were observed in any clinical indicators among the three groups.

Tumor Segmentation and Habitat Subregion Partitioning

The segmentation model achieved mean Dice similarity coefficients of 0.842 and 0.804 on the internal and external validation sets, respectively. When varying the cluster number

k from 2 to 5, the CH index was highest at k = 3; accordingly, each tumor was partitioned into 3 habitat subregions. Segmentation results and habitat subregion partitioning are illustrated in Figure 3.

Backbone Network Comparison

The performance of HMDL models trained with ResNet-18, MobileNet-V2, and DenseNet-121 on the training, internal validation, and external validation sets is shown in Figure 4a. The HMDL (MobileNet-V2) model achieved AUCs of 0.929 and 0.913 on the internal and external validation sets, respectively, higher than the HMDL (ResNet-18) model (0.869 and 0.877) and the HMDL (DenseNet-121) model (0.909 and 0.899). The HMDL (MobileNet-V2) model was selected as the final HMDL model.

Input Strategy Comparison

Regarding input strategy and network dimension, the HMDL (single-plane 2D) model retaining only axial plane input achieved AUCs of 0.885 and 0.872, whereas the HMDL (3D) model replacing the 2D backbone with a 3D backbone showed further decreases to 0.619 and 0.568. Both models had lower AUCs than the HMDL model on both validation sets (Figure 4b). The difference between HMDL and HMDL (single-plane 2D) was not statistically significant on either validation set (P = 0.315, P = 0.565), whereas the difference between HMDL and HMDL (3D) was statistically significant (P = 0.010, P = 0.002).

Ablation Experiment Results

At the imaging sequence level, removal of any single sequence decreased the AUC on both validation cohorts; detailed model performance is shown in Figure 4c. The HMDL (w/

o DCE) model achieved AUCs of 0.873 and 0.855 on the internal and external validation sets, respectively; the difference from the HMDL model was not statistically significant (P = 0.058, P = 0.314). The HMDL (w/o T2FS) model achieved AUCs of 0.909 and 0.879 ; the difference was likewise not statistically significant (P = 0.450, P = 0.519). The HM DL (w/o DWI) model achieved AUCs of 0.702 and 0.676; the difference from the HMDL model was statistically significant (P = 0.015, P = 0.012). Representative Grad-CAM res ults for the three sequences and different slices of the same patient are shown in Figure 5 and Additional files 9 and 10 (Figures S3 and S4). The class activation hotspots for differ ent sequence images were mutually inconsistent, suggesting that each sequence provided complementary discriminative information to the model. For habitat subregion images, th e activation hotspots of each sequence were primarily located within the corresponding su bregion boundaries, indicating that habitat subregion representations provided utilizable d iscriminative information to the model.

At the information source level, the HMDL (w/o habitat), HMDL (w/o radiomics), and H MDL (w/o clinical) models achieved AUCs of 0.885, 0.909, and 0.893 on the internal val idation set and 0.865, 0.896, and 0.891 on the external validation set (Figure 4d), all lowe r than the HMDL model. None of the three models showed statistically significant differe nces from the HMDL model (internal validation set: P = 0.102, P = 0.311, P = 0.145; exte rnal validation set: P = 0.466, P = 0.604, P = 0.600).

Comparison with Clinical-Habitat Radiomics and Clinical Models

Details of features included in the clinical-habitat radiomics and clinical models are provi ded in Additional file 6. The performance of the two models under three machine learnin

g algorithms is shown in Additional files 7 and 8 (Figures S1 and S2). The detailed perfor mance of the HMDL model, clinical-habitat radiomics model, and clinical model across t he three datasets is shown in Figure 6. On the internal validation set, the clinical-habitat r adiomics model achieved an AUC of 0.869, higher than the clinical model (0.736), but th e difference was not statistically significant (P = 0.215); the HMDL model achieved a hig her AUC than both, but the differences were also not statistically significant (P = 0.535, P = 0.079). On the external validation set, the clinical-habitat radiomics model achieved an AUC of 0.841, higher than the clinical model (0.709), but the difference was not statistic ally significant (P = 0.179); the HMDL model achieved a higher AUC than both models, with no statistically significant difference from the clinical-habitat radiomics model (P = 0.357) but a statistically significant difference from the clinical model (P = 0.027).

Discussion

Principal Findings and Model Architecture Strategy

This study proposed and validated a habitat-based multimodal deep learning (HMDL) mo del for noninvasive preoperative discrimination of ALN metastasis status in breast cancer using mpMRI. By jointly optimizing whole-tumor and habitat subregion-level multipara metric MRI images, radiomics features, and clinical features within a unified network end -to-end, the HMDL model synergistically integrated multiparametric MRI imaging, intrat umoral spatial heterogeneity, and clinical information to enable quantitative prediction of ALN metastasis probability. The results showed that the HMDL model achieved an AUC of 0.913 on the independent external validation set, significantly higher than the clinical model (P = 0.027). These findings demonstrate that a multimodal deep learning model gu

ided by intratumoral spatial heterogeneity possesses favorable predictive accuracy and off ers a novel technical approach for preoperative noninvasive ALN assessment in breast ca ncer.

Comparison with Existing Studies

Deep learning technology has achieved notable advances in breast ALN prediction. Wang et al. trained a ResNet50 deep learning network based on T1WI, T2-weighted imaging (T2WI), and DWI sequence MRI, sending results from different sequences to a classifier f or ensemble learning to construct an mpMRI ALN metastasis (ALNM) model; the test set AUC reached 0.996 [7]. Zhang et al. constructed a hierarchical multitask deep learning model, BCALN-Net, in 6,271 patients, first performing automatic segmentation with 3D nnU-Net and then extracting 3D features from DCE, T2WI, and DWI MRI using a 3D De nseNet-121 dual encoder for multitask learning, simultaneously predicting sentinel lymph node metastasis, metastatic burden, and non-sentinel lymph node metastasis, with AUCs of 0.908–0.939, 0.858–0.880, and 0.875–0.911, respectively [13]. Building upon this, the HMDL model proposed in this study introduces habitat subregion images into the multim odal deep learning framework, enabling the model to directly learn intratumoral spatial h eterogeneity information at the subregional level and synergistically model it with radiom ics features and clinical features. The results showed that the HMDL model achieved hig her AUCs than the traditional clinical-habitat radiomics model (internal: 0.929 vs. 0.869; external: 0.913 vs. 0.841) and the clinical model (0.736 and 0.709) on both validation coh orts; the difference from the clinical model was statistically significant on the external val idation set ($P = 0.027$). Although the difference from the clinical-habitat radiomics model did not reach statistical significance, the direction of change was consistent across both c

ohorts, and the lack of significance is more likely attributable to limited statistical power due to small sample sizes. Ablation experiment results showed that removing habitat ima ges from the HMDL (w/o habitat) model decreased the AUC compared with the HMDL model on both the internal and external validation sets, suggesting that habitat images hav e the potential to improve model performance.

It is worth noting that end-to-end deep learning model training always requires a trade-off between sample requirements and spatial information completeness. 2D networks have f ewer parameters and can leverage transfer learning from natural image datasets such as I mageNet to compensate for insufficient samples [14,15,16], but they can only extract 2D image features with limited spatial information utilization. Although 3D networks can dir ectly model 3D data, they are constrained by parameter count and optimization difficulty and are not easily fully trained with limited sample sizes. Given the relatively small samp le size in this study, 2D MobileNet-V2 was adopted as the backbone network, and a 2D t hree-plane input strategy was employed to extract images along the axial, coronal, and sa gittal planes as inputs to the 2D network, preserving the advantages of 2D networks being easy to train and transferable while partially compensating for the missing spatial dimens ion information of single-slice input; similar multi-plane, multi-view analysis strategies h ave been applied and shown value in multiple medical imaging studies [17,18,19]. The in put strategy comparison experiment in this study supports the above consideration: when retaining only axial single-plane input, the model AUCs decreased to 0.885 and 0.872 on the internal and external validation sets, respectively; when switching to a 3D backbone, t hey further decreased to 0.619 and 0.568, both lower than the full model with three-plane

2D input, suggesting that the multi-plane input strategy contributed to maintaining model performance.

Contribution of Multiparametric MRI Sequences

mpMRI can characterize the tumor microenvironment [20] and has been widely used for preoperative assessment of breast ALN metastasis [21,22,23]. Shi et al. systematically co mpared radiomics models based on T1WI, fat-suppressed T2WI, DWI, and DCE in 141 p atients and found that the best-performing sequence combination was DCE combined wit h T2WI and DWI; adding other sequences or removing any one of these three sequences decreased model performance [24]. The DCE, T2FS, and DWI three-sequence combinati on used in this study is essentially consistent with the composition of the best-performing combination; the sequence ablation results in this study are also consistent with the abov e observation: removal of any single sequence decreased the AUC on both validation coh orts, suggesting that DCE, T2FS, and DWI all had a consistent positive contribution trend to the model. Among them, removal of DWI caused the largest AUC decrease, with inter nal and external validation set AUCs dropping to 0.702 and 0.676, respectively, and the d ifference from the original model was statistically significant (P = 0.015, P = 0.012); simi larly, in Shi et al.'s sequence combination comparison, removal of DWI also caused the la rgest AUC decrease on the internal validation set and the second largest on the external v alidation set [24]. Regarding DWI, the degree of water molecule diffusion restriction it re flects is closely related to tissue cell density and extracellular space status, and the tumor stromal component ratio can also be reflected in diffusion quantitative parameters [25]; cl inical studies have also reported that apparent diffusion coefficient (ADC) can be used to predict lymphovascular invasion [26] and axillary lymph node metastasis [27]. The diffus

ion restriction information carried by DWI is relatively more difficult to substitute with ot her sequences, which may explain why removal of DWI caused the most pronounced perf ormance decrease in this study.

Limitations

This study has several limitations. First, this was a retrospective design, and the training d ata were derived from a single center, which may introduce selection bias and spectrum b ias. Second, the sample size was relatively limited, particularly for Center 2, which serve d as the independent external validation set and included only 41 patients; model perform ance requires further validation with multicenter, larger-scale data. Finally, subgroup and robustness analyses were not performed, and the stability of model performance across di fferent molecular subtypes remains to be clarified and should be addressed in future studi es.

Conclusions

In summary, this study performed a direct comparison of the HMDL model with the tradi tional clinical-habitat radiomics model and the clinical model under the same cohort, sam e endpoint, and same reference standard. The HMDL model achieved AUCs of 0.929 and 0.913 on the internal and external validation sets, respectively, and remained stable in in dependent external validation. These results suggest that the HMDL model can provide in dividualized supplementary probabilistic information for preoperative ALN status assess ment. Future work should incorporate decision curve analysis, radiologist reader studies, and prospective multicenter designs to further evaluate its clinical utility.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Affiliated Suqian Hospital of Xuzhou Medical University (approval number: 2023038). Because of its retrospective design, the requirement for informed consent was waived.

Consent for publication

All authors named in this manuscript gave their consent for this publication and take full responsibility for its content.

Availability of data and materials

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Research Project of Jiangsu Provincial Health Commission (Grant No. Z2022023); the Suqian Sci&Tech Program (Grant No. S202314 and No. K202426); the Development Fund of Affiliated Hospital of Xuzhou Medical University (Grant No. XYFY202452); the Suqian Municipal Health Research Project (Grant No. MS202403); and the Jiangsu Provincial Youth Science and Technology Talent Support Program (Grant No. JSTJ-2025-1002).

Clinical trial number: not applicable.

Authors' contributions

Xiancong Gao: Conceptualization, Methodology, Data curation, Writing – original draft. Xun Wang: Data curation, Image annotation. Pei Cai: Image annotation. Jiahai Huang: Data collection. Quanquan Guo: Data curation, Formal analysis. Liyou Huang: Supervision, Writing – review & editing, Funding acquisition. Xiancong Gao and Xun Wang contributed equally to this work as co-first authors. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

REFERENCES

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN esti mates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63. https://doi.org/10.3322/caac.21834.

2. Giuliano AE, Ballman KV, McCall L, et al. Effect of axillary dissection vs no axillary dissection on 10-year overall survival among women with invasive breast cancer and sent inel node metastasis: the ACOSOG Z0011 (Alliance) randomized clinical trial. JAMA. 2 017;318:918–26. https://doi.org/10.1001/jama.2017.11470.

3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77. https://doi.org/10.1148/radiol.2015151169.

4. Wu PQ, Guo FL, Wang J, et al. Development and validation of a dynamic contrast-enhanced magnetic resonance imaging-based habitat and peritumoral radiomic model to predict axillary lymph node metastasis in patients with breast cancer: a retrospective study. Quant Imaging Med Surg. 2024;14:8211–26. https://doi.org/10.21037/qims-24-558.

5. Chen S, Zhang Y, Su Y, et al. Habitat radiomics based on dynamic contrast-enhanced magnetic resonance imaging for assessing axillary lymph node burden in clinical T1-T2 stage breast cancer: a multicenter and interpretable study. J Magn Reson Imaging. 2025;62:751–64. https://doi.org/10.1002/jmri.29796.

6. Lee CF, Lin J, Huang YL, et al. Deep learning-based breast MRI for predicting axillary lymph node metastasis: a systematic review and meta-analysis. Cancer Imaging. 2025;25:44. https://doi.org/10.1186/s40644-025-00863-3.

7. Wang Z, Sun H, Li J, et al. Preoperative prediction of axillary lymph node metastasis in breast cancer using CNN based on multiparametric MRI. J Magn Reson Imaging. 2022;56:700–9. https://doi.org/10.1002/jmri.28082.

8. Song D, Yang F, Zhang Y, et al. Dynamic contrast-enhanced MRI radiomics nomogram for predicting axillary lymph node metastasis in breast cancer. Cancer Imaging. 2022;22:17. https://doi.org/10.1186/s40644-022-00450-w.

9. Zhang W, Yue Z, Ye J, et al. Modulation format identification using the Calinski-Harabasz index. Appl Opt. 2022;61:851–7. https://doi.org/10.1364/ao.448043.

10. Kong L, Qin Y, Li H, et al. Development and validation of a preoperative MRI habita t radiomics model to predict variant histology in bladder cancer. J Magn Reson Imaging. 2025;62:1698–710. https://doi.org/10.1002/jmri.70069.

11. Huang L, Xu L, Wang X, et al. Prediction of EGFR mutations in lung adenocarcinom a via CT images: a comparative study of intratumoral and peritumoral radiomics, deep lea rning, and fusion models. Acad Radiol. 2025;32:4880–92. https://doi.org/10.1016/j.acra.2 025.04.029.

12. Zhang J, Huang W, Li Y, et al. Interpretable machine learning radiomics model predi cts 5-year recurrence-free survival in non-metastatic clear cell renal cell carcinoma: a mul ticenter and retrospective cohort study. Acad Radiol. 2025;32:3788–800. https://doi.org/1 0.1016/j.acra.2025.03.042.

13. Zhang X, Qiu Y, Su Y, et al. An intelligent MRI-based all-in-one diagnostic strategy f or axillary lymph node status in breast cancer. Nat Commun. 2026;17:7864. https://doi.or g/10.1038/s41467-026-74704-7.

14. Hu M, Qian X, Liu S, et al. Structural and diffusion MRI based schizophrenia classifi cation using 2D pretrained and 3D naive convolutional neural networks. Schizophr Res. 2 022;243:330–41. https://doi.org/10.1016/j.schres.2021.06.011.

15. Baheti B, Pati S, Menze B, et al. Leveraging 2D deep learning ImageNet-trained mod els for native 3D medical image analysis. Brainlesion. 2023;13769:68–79. https://doi.org/ 10.1007/978-3-031-33842-7_6.

16. Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med. 2021;128:104115. https://doi.org/10.1016/j.compbiomed.2020.104115.

17. Huang YQ, Chen XB, Cui YF, et al. Enhanced risk stratification for stage II colorectal cancer using deep learning-based CT classifier and pathological markers to optimize adjuvant therapy decision. Ann Oncol. 2025;36:1178–89. https://doi.org/10.1016/j.annonc.2025.05.537.

18. Luo Y, Liu W, Fang T, et al. CARL: cross-aligned representation learning for multi-view lung cancer histology classification. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. Cham: Springer; 2023:358–367. https://doi.org/10.1007/978-3-031-43904-9_35.

19. Wu X, Hui H, Niu M, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. Eur J Radiol. 2020;128:109041. https://doi.org/10.1016/j.ejrad.2020.109041.

20. Hoffmann E, Masthoff M, Kunz WG, et al. Multiparametric MRI for characterization of the tumour microenvironment. Nat Rev Clin Oncol. 2024;21:428–48. https://doi.org/10.1038/s41571-024-00891-1.

21. Zhang Y, Wang Z, Li Z, et al. Multiparametric MRI-based prediction model for sentinel lymph node metastasis in breast cancer: integrating quantitative imaging parameters and reproductive history. Int J Womens Health. 2026;18:1–20. https://doi.org/10.2147/ijwh.s606005.

22. Zhang X, Yang Z, Cui W, et al. Preoperative prediction of axillary sentinel lymph no de burden with multiparametric MRI-based radiomics nomogram in early-stage breast ca ncer. Eur Radiol. 2021;31:5924–39. https://doi.org/10.1007/s00330-020-07674-z.

23. Chen X, Yang Z, Huang R, et al. Development and validation of a point-based scorin g system for predicting axillary lymph node metastasis and disease outcome in breast can cer using clinicopathological and multiparametric MRI features. Cancer Imaging. 2023;2 3:54. https://doi.org/10.1186/s40644-023-00564-9.

24. Shi W, Su Y, Zhang R, et al. Prediction of axillary lymph node metastasis using a ma gnetic resonance imaging radiomics model of invasive breast cancer primary tumor. Canc er Imaging. 2024;24:122. https://doi.org/10.1186/s40644-024-00771-y.

25. Zhang X, Qiu Y, Jiang W, et al. Mean apparent propagator MRI: quantitative assessm ent of tumor-stroma ratio in invasive ductal breast carcinoma. Radiol Imaging Cancer. 20 24;6:e230165. https://doi.org/10.1148/rycan.230165.

26. Mori N, Mugikura S, Takasawa C, et al. Peritumoral apparent diffusion coefficients f or prediction of lymphovascular invasion in clinically node-negative invasive breast canc er. Eur Radiol. 2016;26:331–9. https://doi.org/10.1007/s00330-015-3847-4.

27. Chung J, Youk JH, Kim JA, et al. Role of diffusion-weighted MRI: predicting axillar y lymph node metastases in breast cancer. Acta Radiol. 2014;55:909–16. https://doi.org/1 0.1177/0284185113509094.

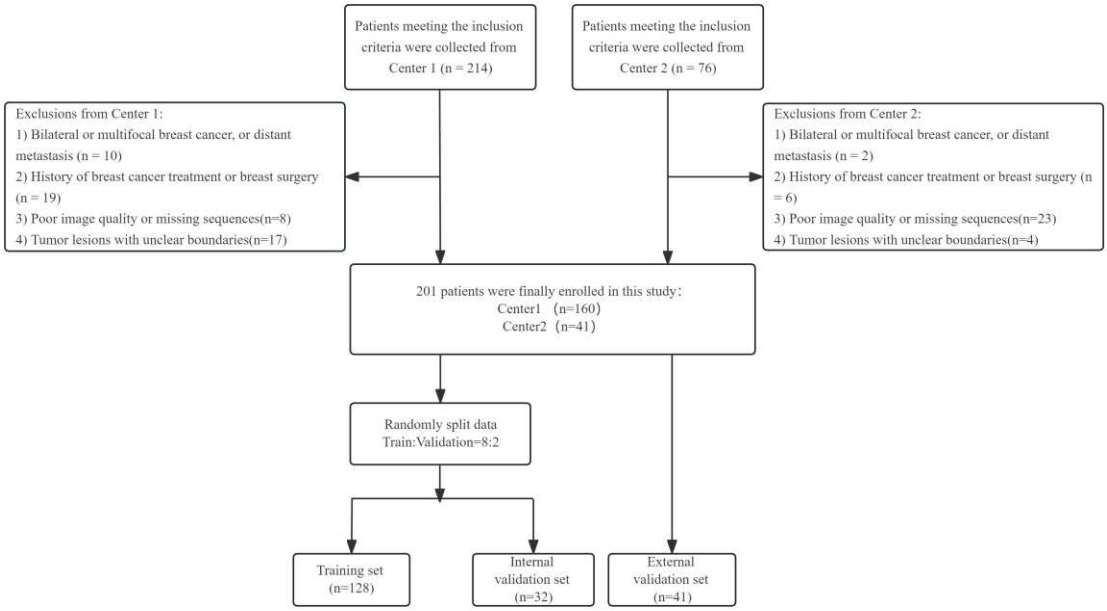


Figure 1. Flowchart of patient enrollment and dataset partitioning.

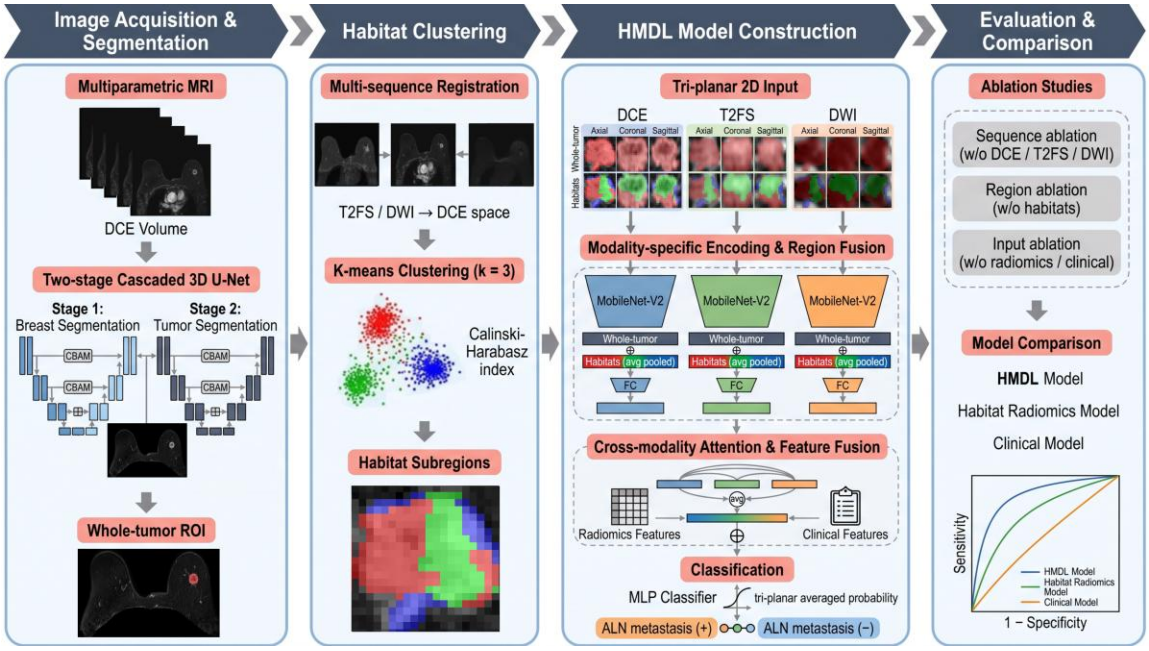


Figure 2. Overall technical workflow of this study.

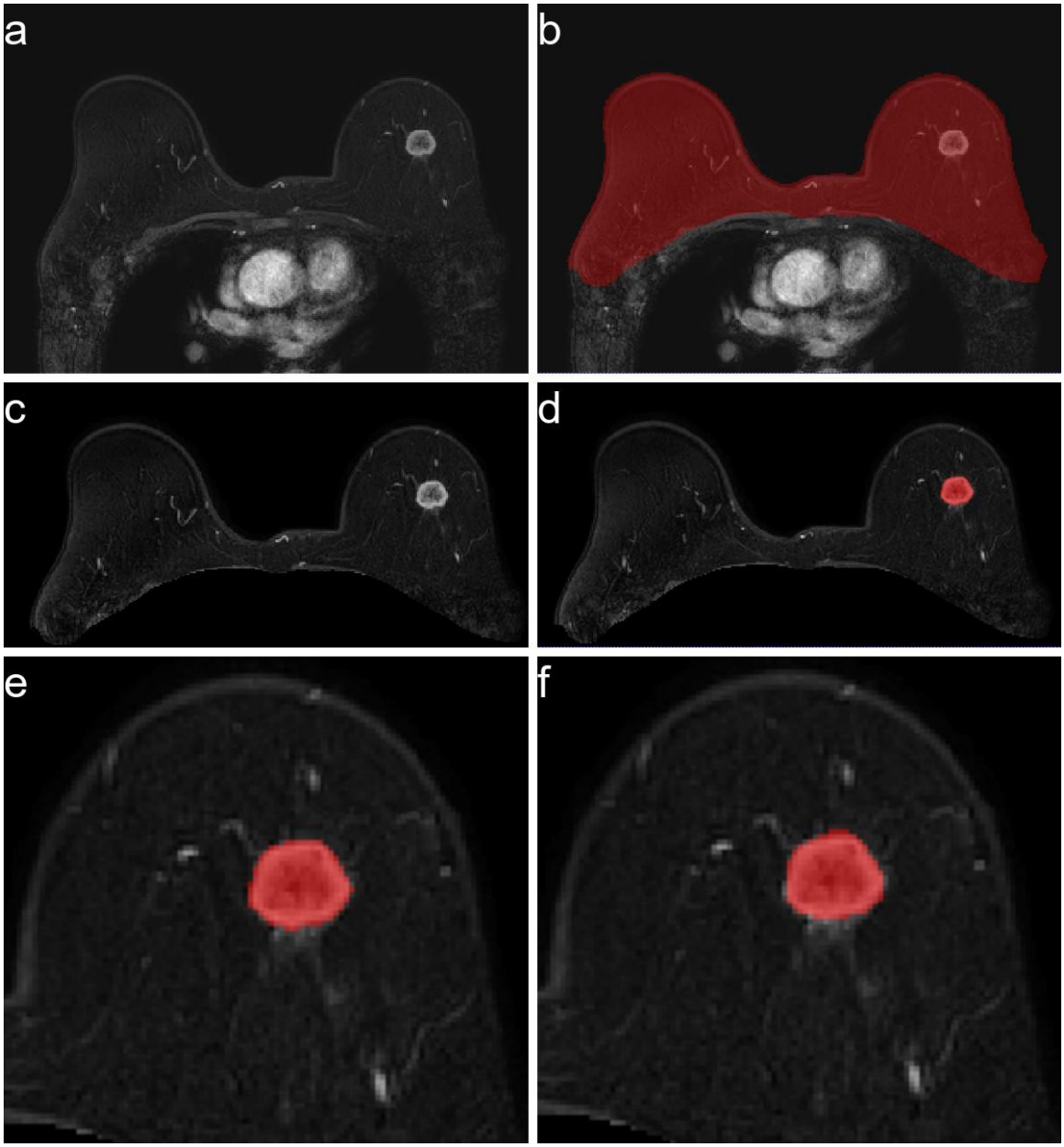


Figure 3. Results of manual and automatic segmentation. (a) Original DCE image; (b) first-stage breast region segmentation result (in red); (c) second-stage model input image cropped based on the first-stage breast mask; (d) breast tumor segmented by the second-stage model (in red); (e) enlarged view of automatic tumor segmentation result; (f) enlarged view of manual segmentation result.

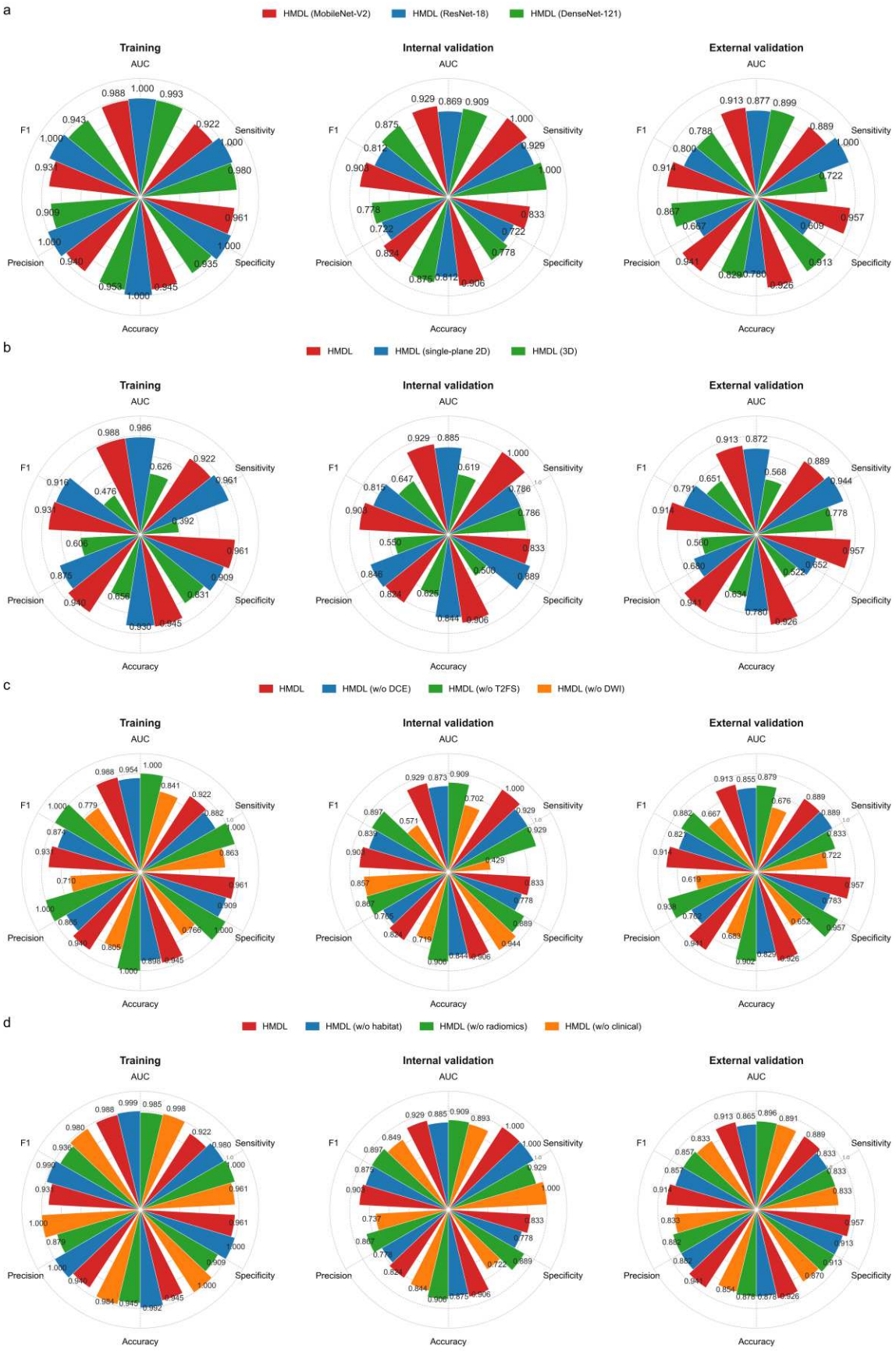

a
HMDL (MobileNet-V2)
HMDL (ResNet-18)
HMDL (DenseNet-121)
Training
Internal validation
External validation
AUC
Sensitivity
Specificity
Accuracy
Precision
F1
b
HMDL
HMDL (single-plane 2D)
HMDL (3D)
c
HMDL
HMDL (w/o DCE)
HMDL (w/o T2FS)
HMDL (w/o DWI)
d
HMDL
HMDL (w/o habitat)
HMDL (w/o radiomics)
HMDL (w/o clinical)

Figure 4. Performance of different HMDL models on the training, internal validation, and external validation sets. (a) Backbone networks: HMDL (MobileNet-V2), HMDL (ResNet-18), and HMDL (DenseNet-121) models; (b) Input strategies: HMDL, HMDL (single-plane 2D), and HMDL (3D) models; (c) Sequence ablation: HMDL (w/o DCE), HMDL (w/o T2FS), and HMDL (w/o DWI) models; (d) Information source ablation: HMDL (w/o habitat), HMDL (w/o radiomics), and HMDL (w/o clinical) models.

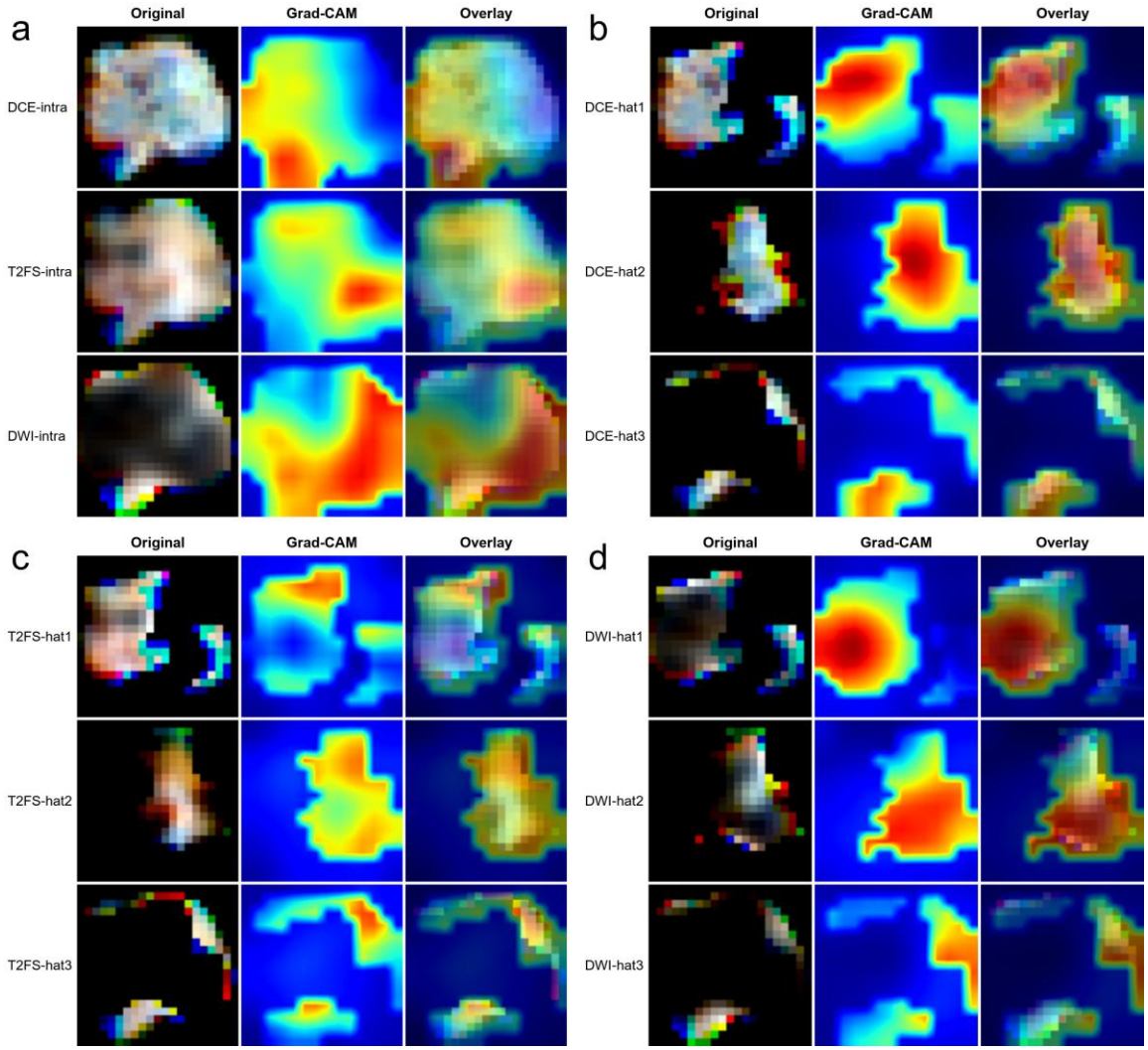


Figure 5. Representative gradient-weighted class activation mapping (Grad-CAM) visualization results of the HMDL model for DCE, T2FS, and DWI sequence images. For each i

nput image, from left to right: original image, Grad-CAM heatmap, and overlay. (a) Who le-tumor images of DCE, T2FS, and DWI sequences in the axial plane; (b-d) axial-plane i mages of three habitat subregions (hat1–hat3) for DCE, T2FS, and DWI sequences, respe ctively.

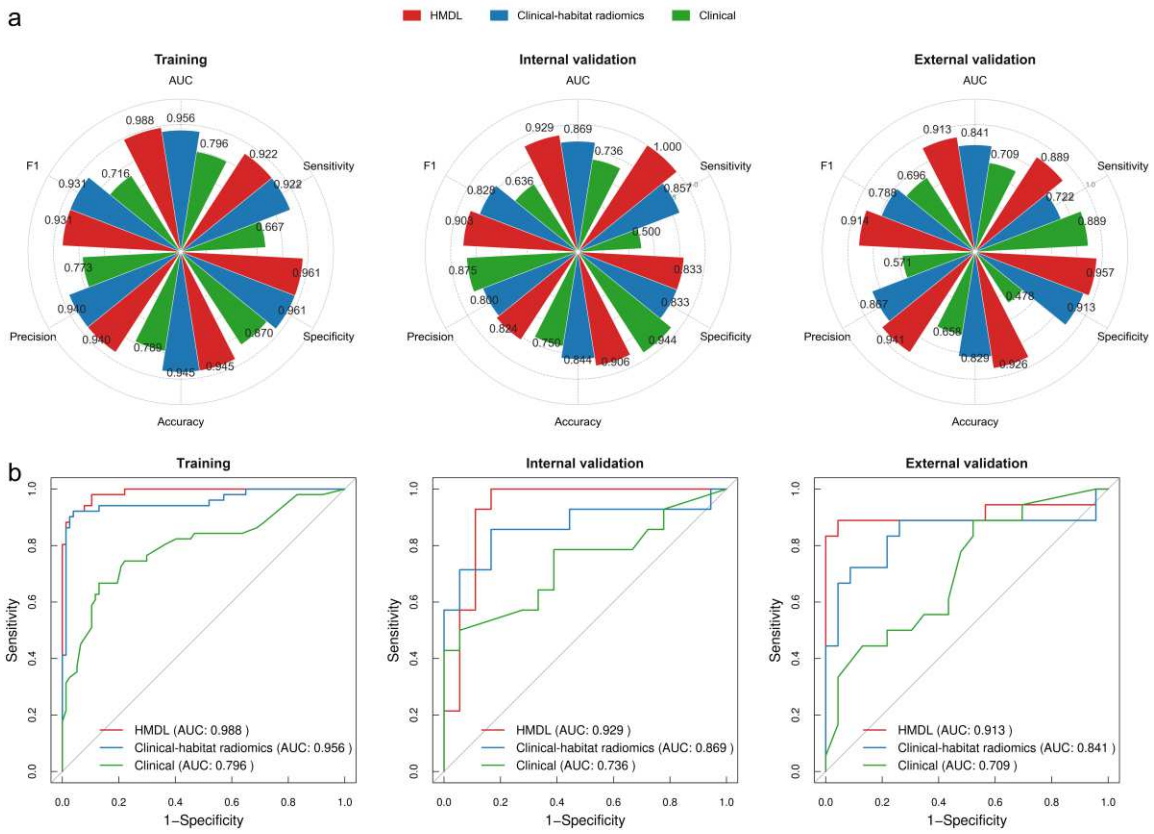


Figure 6. Performance of the three models on the training, internal validation, and externa l validation sets. (a) Detailed performance of the three models, (b) ROC curves of the thre e models.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Additionalfiles.docx