Post-print: Efficacy of Ultrasound Habitat Imaging for Benign-Malignant Differentiation of Breast Phyllodes Tumors

Authors: Xie Danling, Liu Boya, Xiaoguang Li, Hanwei Wang, Ma Qiang, Fang Jingqin, Wang Shunan, Wang Shunan

Date: 2025-11-17T00:00:00+00:00

Abstract

Background: Phyllodes tumors of the breast (PTB) exhibit significant differences in surgical strategies and risks of recurrence and metastasis between benign and malignant forms. Preoperative differentiation between benign and malignant PTB is crucial for treatment decision-making. Conventional ultrasound diagnosis has limitations, and the efficacy of ultrasound habitat imaging in differentiating benign and malignant PTB has not been systematically elucidated.

Objective: To evaluate the efficacy of ultrasound habitat imaging in differentiating benign and malignant PTB.

Methods: Clinical and ultrasound imaging data from 102 patients with surgically pathologically confirmed PTB at Daping Hospital of Army Medical University between September 2014 and June 2024 were retrospectively analyzed. Based on pathological diagnosis, patients were divided into a benign group (54 cases) and a borderline/malignant group (48 cases, including 30 borderline and 18 malignant cases). Ultrasound images were recorded and ITK-SNAP software was used to manually delineate the tumor region of interest (ROI). The ROI was partitioned into 3 habitat subregions via K-means clustering, and habitat features were extracted using PyRadiomics. Random forest (RF) screening was employed to retain optimal features and calculate the habitat radiomics score (Hab-score) to construct a habitat model. Conventional ultrasound variables with statistically significant differences in univariate analysis were included to construct a conventional ultrasound model. Conventional ultrasound features and Hab-score were incorporated to construct a combined model. Receiver operating characteristic (ROC) curves and Delong test were utilized to compare the diagnostic efficacy of each model, and decision curve analysis (DCA) was employed to evaluate the clinical applicability of the models.

Results: Significant differences were observed between the two groups in maximum diameter, internal echo, boundary, and cystic degeneration (P<0.05). The aforementioned 4 ultrasound variables were included to construct the conventional ultrasound model. Following RF screening, 7 habitat features (3 first-order features and 4 texture features) were retained to calculate the Hab-score and construct the habitat model. Based on the 4 conventional ultrasound variables, Hab-score was further incorporated to construct the combined model. The AUCs of the conventional ultrasound model, habitat model, and combined model for differentiating benign and malignant phyllodes tumors were 0.718, 0.725, and 0.799, respectively. Delong test results demonstrated that the AUC of the combined model was superior to those of the conventional ultrasound model and habitat model (P<0.05). DCA curve analysis revealed that the combined model exhibited the highest clinical benefit in the threshold range of 0.4-0.9 for differentiating benign and malignant PTB.

Conclusion: Ultrasound habitat imaging can be effectively applied to differentiate benign and malignant phyllodes tumors of the breast. Integration with conventional ultrasound can further enhance diagnostic efficacy and reduce the risk of missed diagnosis and misdiagnosis associated with single-technology approaches, demonstrating potential clinical application value.

Full Text

Introduction

Phyllodes tumors of the breast (PTB) are rare fibroepithelial neoplasms. According to the 2019 WHO 5th edition classification of breast tumors [1], PTB are categorized as benign, borderline, or malignant, with distinct surgical approaches for each type: benign tumors are preferably managed with excisional biopsy or local wide excision, while borderline/malignant tumors require local extended excision with optimal negative margin width $ $1 cm; patients in whom negative margins cannot be achieved should undergo total mastectomy. Recurrence rates vary significantly across pathological subtypes: <5% for benign, 5%-15% for borderline, and 15%-30% for malignant PTB, with the latter also showing higher propensity for distant metastasis and poorer prognosis correlating with malignant grade [2]. Therefore, accurate preoperative differentiation of PTB malignancy is of critical clinical value for optimizing surgical strategy and improving patient outcomes.

Ultrasound examination is widely used in breast tumor imaging assessment due to its convenience and real-time capabilities. However, the sonographic appearance of PTB substantially overlaps with that of fibroadenomas, making diagnosis highly operator-dependent and limiting its clinical utility [3]. In recent years, habitat imaging [4] has emerged as a novel approach for analyzing tumor microenvironments based on imaging data. By employing spatial segmentation techniques to partition tumors into subregions with distinct biological characteristics, habitat imaging can identify different cell populations and infer tumor

heterogeneity, offering new avenues to overcome the limitations of conventional imaging. While habitat imaging has been predominantly applied in CT and MRI studies [5-6], its application in ultrasound remains limited [7], particularly for PTB. This study aims to differentiate benign from malignant PTB using ultrasound habitat imaging, providing quantifiable evidence for precise preoperative diagnosis and informing clinical treatment decisions and prognostic assessment.

1 Subjects and Methods

1.1 Study Subjects

We retrospectively analyzed 102 patients with pathologically confirmed PTB who underwent surgery at Daping Hospital, Army Medical University between September 2014 and June 2024. Based on pathological results, patients were divided into a benign group (n=54) and a borderline/malignant group (n=48, including 30 borderline and 18 malignant cases). Inclusion criteria were: (1) primary, solitary lesion; (2) complete conventional ultrasound imaging data; and (3) postoperative pathological confirmation of PTB with clear pathological grading. Exclusion criteria were: (1) pregnancy or lactation; (2) prior surgery, chemotherapy, or immunotherapy; and (3) incomplete clinical data. This study was approved by the Ethics Committee of Army Medical Center [Approval No.: (2024) 359]. As a retrospective study without intervention in clinical treatment or prognosis, the requirement for informed consent was waived.

1.2 Methods

1.2.2 Ultrasound Examination and Image Analysis (1) Image Acquisition: Ultrasound examinations were performed using either a Mindray DC-8 or GE Logiq E9 color Doppler ultrasound diagnostic system with L11-3U or ML6-15 probes at 3-12 MHz frequency. Depth was preset to 2-5 cm and adjusted to fully display the lesion, with focal zone positioned at the lesion level and $ $2 focal points. Overall gain was adjusted until subcutaneous fat appeared as medium-level echoes, and color gain was tuned to just below the threshold for color overflow. Patients were positioned supine with hands raised to fully expose both breasts and axillary regions. Transverse, longitudinal, and coronal scans were performed across all breast quadrants. Two-dimensional ultrasound was performed first, and images were stored in DICOM format in the Picture Archiving and Communication Systems (PACS), followed by color Doppler flow examination of the target breast mass.

(2) Image Analysis: Two physicians with at least 5 years of breast ultrasound experience independently reviewed the images without knowledge of postoperative pathology; disagreements were resolved through consultation. Two-dimensional images were described using the Breast Imaging Reporting and Data System (BI-RADS) ultrasound lexicon. Morphological features assessed included maximum tumor diameter, growth pattern (parallel or non-parallel), boundary (clear or unclear), shape (irregular or regular), internal echo (homoge-

neous or heterogeneous), cystic change (absent or present), calcification (present or absent), and posterior acoustic features (shadowing, no change, or enhancement). Blood flow was graded using the Adler semi-quantitative method [8] and classified as grade 0-I (hypovascular) or grade II-III (hypervascular).

1.2.3 Ultrasound Image Segmentation and Habitat Feature Extraction (1) Ultrasound Image Segmentation: Two ultrasound physicians with over 5 years of breast ultrasound diagnostic experience performed the segmentation. They browsed ultrasound images, selected the plane with good image quality showing the maximum diameter, and manually delineated the region of interest (ROI) on grayscale ultrasound images using the open-source imaging platform ITK-SNAP (http://www.itksnap.org). The first physician performed initial segmentation independently, followed by the second physician. Disagreements were resolved through consultation, with ROIs traced along tumor boundaries.

(2) Habitat Partitioning: Using pixel grayscale values within the ROI as input, unsupervised K-means clustering was performed with the Calinski-Harabasz (CH) index to determine the optimal number of clusters (k=3). The tumor interior was divided into three grayscale-homogeneous subregions, generating pseudocolor label maps to visually present intratumor heterogeneity (ITH) in PTB. Examples of ROI delineation and habitat partitioning are shown in [Figure 1: see original paper].

(3) Ultrasound Habitat Feature Extraction: Within each subregion, radiomics features were extracted using PyRadiomics in Python. First-order statistical features included mean, standard deviation, and pixel area. Higher-order texture features were extracted from the gray-level co-occurrence matrix (distance=1, angle=0°) including contrast, dissimilarity, homogeneity, angular second moment (ASM), and correlation. If a subregion contained <10 pixels, all texture features for that subregion were assigned zero values to avoid noise interference.

1.2.4 Feature Selection for Conventional Ultrasound and Habitat Features The intraclass correlation coefficient (ICC) was used to assess inter-observer consistency in feature extraction; features with ICC<0.75 were excluded as unreliable. All habitat features underwent Z-score normalization. Conventional ultrasound features were screened using univariate analysis to identify variables with statistically significant differences. Habitat features were selected using random forest (RF) algorithm with a selection criterion of feature importance $ $0.05 to retain core habitat features with significant model contribution. Based on these selected core habitat features, a habitat score (Hab-score) was calculated as the core input variable for the habitat model.

The Hab-score calculation process was as follows: First, the 7 habitat features were Z-score standardized (based on mean and standard deviation across all samples to eliminate dimensional differences). Then, weighted summation was

performed using "standardized feature value × corresponding weight" to quantify habitat features into Hab-score. The formula was: Hab-score = Σ(i=1 to 7)(Xi,std × Wi), where Xi,std represents the standardized value of the ith feature and Wi represents its RF weight coefficient.

1.2.5 Model Construction After feature selection, three types of models were constructed using RF methodology based on variable source: (1) Conventional ultrasound model (including variables with statistically significant differences in univariate analysis of ultrasound features); (2) Habitat model (including variables: Hab-score calculated from core habitat features selected by RF with feature importance $ $0.05); and (3) Combined model (including variables: ultrasound features with statistically significant differences in univariate analysis plus Hab-score).

1.3 Statistical Methods

Statistical analysis was performed using Python (version 3.11.8). Continuous variables with normal distribution were expressed as ($\bar{x}\pm s$) and compared using independent two-sample t-tests; non-normally distributed continuous variables were expressed as M(P25, P75) and compared using Mann-Whitney U test. Categorical data were expressed as proportions and compared using chi-square test. Receiver operating characteristic (ROC) curves were plotted to evaluate model predictive performance, with DeLong test used to compare AUC differences between models. Decision curve analysis (DCA) was employed to assess clinical utility. $P<0.05$ was considered statistically significant.

2 Results

2.1 Comparison of General Data and Ultrasound Features Between Groups

A total of 102 PTB patients were included, all female, with mean age (45.7$±$12.1) years: 54 in the benign group and 48 in the borderline/malignant group. Statistically significant differences were observed between groups in maximum tumor diameter, boundary, internal echo, and cystic change ($P<0.05$), as shown in .

2.2 Habitat Feature Selection and Model Establishment

The conventional ultrasound model was constructed using 4 ultrasound variables with significant differences in univariate analysis (maximum tumor diameter, boundary, internal echo, cystic change). From the tumor ROI, 3 subregions (habitat_1, habitat_2, habitat_3) were extracted, yielding 24 habitat features. After RF screening, 7 habitat features (3 first-order features, 4 texture features) were retained for Hab-score calculation and habitat model construction. Among these, habitat_1_{homogeneity} contributed the highest weight

(+0.081), representing the most discriminative habitat feature, followed by habitat_3_{mean} (+0.064) and habitat_2_{mean} (+0.064), as shown in .

2.3 Performance Evaluation of Each Model

ROC curves were plotted for the conventional ultrasound model, habitat model, and combined model to differentiate benign and malignant PTB. The conventional ultrasound model achieved an AUC of 0.718 (95%CI=0.610-0.816) with accuracy 67.7%, sensitivity 58.3%, and specificity 75.9%. The habitat model achieved an AUC of 0.725 (95%CI=0.620-0.814) with accuracy 61.8%, sensitivity 68.8%, and specificity 55.6%. The combined model achieved an AUC of 0.799 (95%CI=0.712-0.876) with accuracy 70.6%, sensitivity 62.5%, and specificity 77.8%, as shown in and [Figure 3: see original paper]. DeLong test results indicated that the combined model' s AUC was significantly higher than those of the conventional ultrasound model and habitat model (P<0.05). DCA curve analysis demonstrated that the combined model provided the highest clinical net benefit for PTB differentiation within the threshold range of 0.4-0.9, as shown in [Figure 4: see original paper].

Comparison of optimal habitat features and overall habitat features between the benign and borderline/malignant groups revealed that median values of habitat_2_{mean} and habitat_3_{std} were higher in the benign group, while median values of habitat_1_{homogeneity}, habitat_3_{mean}, habitat_2_{correlation}, habitat_1_{correlation}, and habitat_1_{ASM} were lower compared to the borderline/malignant group (P<0.05). The median Hab-score was also lower in the benign group than in the borderline/malignant group (P<0.05), as shown in [Figure 2: see original paper]. The combined model was constructed by incorporating Hab-score into the 4 conventional ultrasound variables.

Discussion

In recent years, conventional ultrasound, as one of the most widely used imaging modalities in clinical practice, offers real-time and non-invasive advantages, yet its potential for quantifying PTB heterogeneity remains underexplored. Habitat imaging, an innovative technique for quantifying spatial heterogeneity in tumor microenvironments, has demonstrated effectiveness in CT and MRI for assisting tumor benign-malignant differentiation and biological behavior assessment, garnering significant attention in diagnostic imaging. However, the application of ultrasound habitat imaging in PTB differentiation has not been systematically investigated. This study established a combined model based on conventional ultrasound and habitat features to improve ultrasound diagnostic efficacy for PTB, aiming to provide a non-invasive assessment approach for preoperative clinical decision-making.

Our univariate analysis of conventional ultrasound features identified statistically significant differences in maximum tumor diameter, internal echo, bound-

ary, and cystic change (P<0.05). Malignant breast tumors exhibit faster growth rates and larger diameters than benign lesions, with rapid proliferation causing internal structural disorganization manifested as necrosis, hemorrhage, or fibrosis [9], which appears as heterogeneous echotexture on ultrasound. Benign PTB typically have intact fibrous capsules, whereas malignant PTB show stromal overgrowth, significant nuclear atypia, and mitotic figures $ $10/10 high-power fields (HPF) [10], with invasive growth into surrounding breast ducts, lobules, or glandular tissue that can breach the capsule, resulting in unclear tumor boundaries [11]. According to Cao et al. [12], approximately 88.89% of benign PTB had clear boundaries with only 11.11% showing blurred margins, consistent with our findings. Additionally, rapid proliferation in malignant PTB leads to imbalanced internal blood supply, predisposing to micro-necrosis and hemorrhage, which manifest as internal cystic changes on ultrasound. LI et al. [13] used logistic regression to evaluate ultrasound features for predicting PTB malignancy in 79 cases, identifying significant differences in age, lesion size, morphology, internal echo, liquefaction, and blood flow (P<0.05). These results differ somewhat from our univariate analysis, possibly due to our larger sample size (102 vs. 79 cases) requiring multi-center validation. Regarding blood flow, our study showed richer vascular signals in the malignant group (higher proportion of grade II-III), likely attributable to stromal overgrowth in borderline/malignant PTB causing cellular disarray, irregular internal architecture, disruption of existing vascular beds, and induction of numerous immature neovessels, resulting in abundant color Doppler signals, though further research is needed to precisely characterize tumor vascularity.

This study partitioned tumor ROIs into three subregions using K-means clustering, extracting ultrasound habitat features to establish a combined model with conventional ultrasound features. The retained features were habitat_1_{homogeneity}, habitat_3_{mean}, habitat_2_{mean}, habitat_3_{std}, habitat_2_{correlation}, habitat_1_{correlation}, and habitat_1_{ASM}. Further analysis of feature distribution revealed that habitat_1_{ASM} quantifies texture uniformity and concentration, with higher values indicating greater texture consistency and regularity—characteristics associated with relatively slower-growing benign tumors with more mature structural differentiation and orderly arrangement [14]. Additionally, increased habitat_3_{std} reflects higher grayscale dispersion within malignant lesions, likely due to coexistence of malignant cells with internal necrosis and hemorrhage components creating more pronounced grayscale differences. Elevated habitat_3_{mean} and habitat_2_{mean} indicate higher grayscale levels in malignant groups, correlating with enhanced signals from intensified malignant cell proliferation. Increased levels of habitat_1_{homogeneity}, habitat_2_{correlation}, and habitat_1_{correlation} reflect stronger spatial correlations of grayscale values in malignant tumors, possibly resulting from abnormal tumor cell proliferation forming pathological structures (such as tumor nests and aberrant vascular networks) that exhibit abnormal but regular vascular distributions, leading to more uniform image textures. These habitat

features are all associated with enhanced tissue heterogeneity resulting from disordered malignant cell proliferation and invasive growth, demonstrating that feature differences across habitat subregions can reflect intratumoral heterogeneity [15]. ZHANG et al. [16] retrospectively analyzed multiparametric MRI images of 76 breast cancer patients, using habitat features to quantitatively visualize differences in vascular distribution and cellularity within breast cancer lesions, finding that habitat imaging could identify breast cancer heterogeneity—consistent with our results. However, that study further analyzed differences in volume proportions across habitats for breast cancers with different HER2 expression levels, whereas ultrasound habitat analysis correlated with pathological markers remains to be investigated.

Comparing the efficacy of conventional ultrasound, habitat, and combined models for PTB differentiation, the combined model performed best, surpassing both conventional ultrasound and habitat models. DeLong test further confirmed statistically significant differences in AUC between the combined model and both conventional ultrasound and habitat models ($P<0.05$), while no significant difference existed between the conventional ultrasound and habitat models ($P>0.05$), suggesting limited discriminative value of either approach alone and necessitating multi-feature integration for breakthrough diagnostic improvement. XU et al. [17] used habitat MRI analysis to differentiate triple-negative from non-triple-negative breast cancer in 142 patients, achieving an AUC of 0.951 for their combined model—substantially higher than our combined model' s AUC of 0.799. This discrepancy likely reflects inherent differences in imaging modalities: MRI, as a multiparametric functional imaging technique, can quantify physiological states of tumor microenvironments based on microcirculatory perfusion and water molecule diffusion characteristics, making MRI habitat partitions more closely aligned with essential tumor biological behaviors. Ultrasound imaging, based on sound wave reflection and scattering principles, can only indirectly reflect tissue density differences and structural arrangement through grayscale texture features, while color Doppler parameters only capture blood flow signals from vessels >0.1mm diameter, unable to directly assess microvascular angiogenesis or cellular-molecular features, resulting in relatively lower biological specificity of ultrasound habitat features compared to MRI. Additionally, the conventional ultrasound model' s sensitivity of only 58.3% reflects missed diagnosis risk for borderline/malignant PTB, whereas the habitat model' s sensitivity increased to 68.8%, suggesting that quantifying intratumoral habitat features can more sensitively capture elevated heterogeneity in malignant tumors. However, the habitat model' s specificity of only 55.6% indicates that some benign PTB may exhibit habitat features overlapping with malignant PTB due to internal changes such as cystic degeneration. Therefore, the complementary advantages of the two single models highlight the significant benefits of feature integration in the combined model. Yu et al. [18] constructed a training set and test set AUC of 0.825 and 0.818, respectively, for an ultrasound + texture feature model based on 298 PTB patients—slightly higher than our combined model' s AUC (0.799), likely attributable to their substantially larger

sample size reducing random error and enhancing the model' s ability to capture independent predictors. Additionally, their study used internal validation whereas ours did not. Despite different research strategies, our combined model AUC of 0.799 closely approaches their results, reflecting good discriminative performance.

From a clinical perspective, when breast masses suspected as PTB are detected by palpation or conventional ultrasound but cannot be clearly distinguished as benign or malignant, ultrasound habitat imaging can be introduced for combined diagnosis alongside conventional ultrasound assessment, thereby overcoming technical limitations of traditional ultrasound in tumor heterogeneity evaluation. However, this study has several limitations: (1) The sample size was limited and the study was single-center retrospective; multi-center studies with larger sample sizes are needed to further validate model generalizability. (2) Our ultrasound images were two-dimensional, lacking three-dimensional perspective; constructing 3D models could enrich habitat imaging features and more comprehensively reflect PTB internal heterogeneity. (3) Our ultrasound images were acquired using two different ultrasound systems (Mindray DC-8 and GE Logiq E9). Although core parameters such as probe frequency (3-12 MHz) and depth adjustment were standardized, inherent differences exist between brands in imaging resolution, grayscale processing algorithms, and color gain optimization logic. Additionally, subtle parameter adjustments by different operators during image acquisition may introduce bias.

In summary, this study effectively applied ultrasound habitat imaging technology to differentiate benign and malignant phyllodes tumors of the breast. Combined with conventional ultrasound features, diagnostic efficacy was further improved, reducing missed diagnosis and misdiagnosis risks associated with single-technical approaches, and providing a novel non-invasive technical pathway for ultrasound assessment of PTB heterogeneity.

Author Contributions: Xie Danling was responsible for study implementation and manuscript writing. Xie Danling and Liu Boya collected and organized ultrasound data, performed statistical analysis, created figures and tables, and revised the manuscript. Li Xiaoguang participated in experimental design. Wang Hanwei participated in statistical analysis. Ma Qiang provided pathological data. Fang Jingqin participated in research protocol revision and manuscript modification. Wang Shunan was responsible for experimental design, quality control and review of the manuscript, and supervision.

Conflict of Interest: The authors declare no conflict of interest.

Note: Figure translations are in progress. See original paper for figures.