Differential diagnosis between small breast phyllodes tumors and fibroadenomas using artificial intelligence and ultrasound data

Sihua Niu[1], Jianhua Huang[2], Jia Li[3], Xueling Liu[4], Dan Wang[4], Yingyan Wang[3], Huiming Shen[3], Min Qi[3], Yi Xiao[2], Mengyao Guan[2], Diancheng Li[1], Feifei Liu[1], Xiuming Wang[1], Yu Xiong[1], Siqi Gao[1], Xue Wang[1], Ping Yu[1], Jia'an Zhu[1]

[1]Department of Ultrasound, Peking University People's Hospital, Beijing, China; [2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China; [3]Department of Ultrasound, Zhongda Hospital Southeast University, Nanjing, China; [4]Department of Ultrasound, The First Affiliated Hospital of Guangxi University of Traditional Chinese Medicine, Nanning, China

Correspondence to: Jiaan Zhu. Department of Ultrasound, Peking University People's Hospital, 11 Xizhimen South Street, Xicheng District, Beijing 100044, China. Email: zhujiaan@pkuph.edu.cn.

Background: It is challenging to differentiate between phyllodes tumors (PTs) and fibroadenomas (FAs). Artificial intelligence (AI) can provide quantitative information regarding the morphology and textural features of lesions. This study attempted to use AI to evaluate the ultrasonic images of PTs and FAs and to explore the diagnostic performance of AI features in the differential diagnosis of PTs and FAs.

Methods: A total of 40 PTs and 290 FAs <5 cm in maximum diameter found in female patients were retrospectively analyzed. All tumors were segmented by doctors, and the features of the lesions were collated, including circularity, height-to-width ratio, margin spicules, margin coarseness (MC), margin indistinctness, margin lobulation (ML), internal calcification, angle between the long axis of the lesion and skin, energy, grey entropy, and grey mean. The differences between PTs and FAs were analyzed, and the diagnostic performance of AI features in the differential diagnosis of PTs and FAs was evaluated.

Results: Statistically significant differences ($P<0.05$) were found in the height-to-width ratio, ML, energy, and grey entropy between the PTs and FAs. Receiver operating characteristic (ROC) curve analysis of single features showed that the area under the curve [(AUC) 0.759] of grey entropy was the largest among the four features with statistically significant differences, and the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were 0.925, 0.459, 0.978, and 0.190, respectively. When considering the combinations of the features, the combination of height-to-width ratio, margin indistinctness, ML, energy, grey entropy, and internal calcification was the most optimal of the combinations of features with an AUC of 0.868, and a sensitivity, specificity, PPV, and NPV of 0.734, 0.900, 0.982, and 0.316, respectively.

Conclusions: Quantitative analysis of AI can identify subtle differences in the morphology and textural features between small PTs and FAs. Comprehensive consideration of multiple features is important for the differential diagnosis of PTs and FAs.

Keywords: Artificial intelligence (AI); breast; ultrasound; phyllodes tumor (PT); fibroadenoma

Introduction

Phyllodes tumors (PTs) and fibroadenomas (FAs) are both biphasic fibroepithelial tumors with similar histological characteristics. Occurring mostly in young women, FAs are the most common benign breast tumors. Comparatively, PTs are rare, and account for ~0.3–1% of breast tumors (1). The age of onset for PTs is later than that of FAs, and the volume of PTs is often greater than that of FAs. PTs can reach an average size of 6.4 cm (2) and have abundant stromal cells and typically leaf-like structures. According to histological characteristics such as the number of stromal cells, mitotic figures, cell atypia, overgrowth of stroma, and the nature of the tumor border, PTs can be divided into 3 pathological types: benign, borderline, and malignant (3). Benign PTs have a large quantity of stromal cells, while in malignant PTs, increased mitotic figures of stromal cells and marginal infiltration are commonly seen. The characteristics of borderline PTs are somewhere between those of benign and malignant PTs (3). After surgery, recurrence may occur in benign, borderline, and malignant PTs at rates of 14–17%, 14–25%, and 23–30%, respectively (4,5), with 22% of patients being reported experiencing metastasis or death (5). Therefore, the correct diagnosis of PTs is critical for the treatment and prognosis of patients.

Mammography, magnetic resonance imaging (MRI), and ultrasonography are commonly used to diagnose breast lesions (6,7). Among these methods, ultrasonography is more commonly used in China for diagnosing breast lesions through the observation of features such as size, shape, orientation, margin, and echo pattern. Both PTs and FAs can present as regular or irregular solid masses in images, and it is challenging to differentiate PTs from FAs using radiologic features (2,3,8). Both PTs and FAs can be classified into the Breast Imaging Reporting and Data System (BI-RADS) 3 or 4A categories according to the ACR BI-RADS® Atlas Fifth Edition, which reflects the risk stratification of breast lesions. BI-RADS 3 lesions are probably benign, and BI-RADS 4A lesions have low suspicion for malignancy (9). According to the literature, 50% of PTs are initially misdiagnosed as FAs (2), and some cases eventually require biopsy. However, due to the different biological behaviors of PTs and FAs, PTs require local extended resection (tumors should be resected with ≥1 cm margins), while FAs can be treated non-surgically or by simple enucleation (only the tumor is resected). Therefore, differential diagnosis is particularly important when considering treatment.

Deep learning and other technologies of artificial intelligence (AI) can accurately identify lesions in images, calculate and obtain quantitative information on the morphology and textural features of lesions, and thus provide a basis for differentiating breast diseases (10,11).

Presently, studies on AI for differentiating PTs from FAs are limited and have mainly focused on MRI and ultrasonic images (12-14). Research has shown that the textural features on MRI T2-weighted short-tau inversion recovery (T2W-STIR) have higher diagnostic performance compared with clinical and conventional MRI features for distinguishing between PTs and FAs (12). Deep learning software has the capability to differentiate between PTs and FAs with good diagnostic accuracy and high negative predictive value (NPV) (14). Previous studies have focused on other imaging modalities or comparison with doctors' diagnostic efficacy, and AI studies on ultrasound images have only compared differences between PTs, complex FAs, and simple FAs (13). Thus far, however, further evaluation of the diagnostic performance of each ultrasonic feature and consideration of the influence of lesion size on the differential diagnosis has not been undertaken.

An overlap exists between the size and ultrasonic features of PTs and FAs. A large volume can indicate the possibility of PTs, whereas for small PTs, it is difficult to make an ultrasonic diagnosis with the naked eye. Thus, in our study, we attempted to use AI to compare differences between small PTs and FAs, determine the morphology and textural features with diagnostic value, and evaluate the diagnostic performance of those features.

Methods

Participants

All participants were patients at the Peking University People's Hospital, Southeast University Zhongda Hospital, and the First Affiliated Hospital of Guangxi University of Traditional Chinese Medicine. This study was approved by the ethics board of the Peking University People's Hospital (No. 2018PHB179-01), and informed consent was provided by all participants. We reviewed all of the original ultrasonic images of the participants diagnosed with PTs or FAs from January to December 2019. The inclusion criteria were as follows: (I) the tumors were able to be shown in a linear-array probe, and considering that the width of the probes used in our research was ≤5 cm, lesions <5 cm were regarded as small tumors and included in our study; (II) there were no measurement labels in the greyscale images; (III) all cases were surgically resected and pathologically

2054

Niu et al. Difference between phyllodes tumors and fibroadenomas
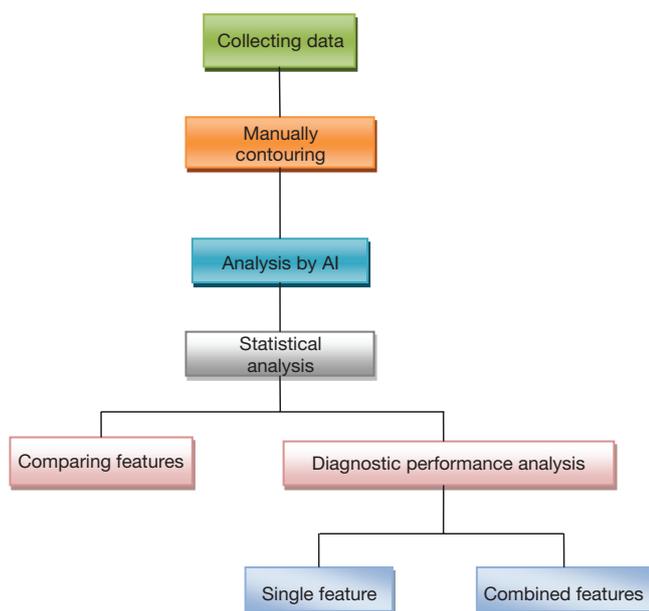
Figure 1 Overview of the methods and procedures. AI, artificial intelligence.

diagnosed; and (IV) clear definitions of benign, malignant, and borderline in the pathological diagnosis of PTs were able to be obtained. The exclusion criteria were as follows: (I) tumors were shown in color Doppler flow images; (II) there were measurement labels in greyscale images; (III) the tumors could not be displayed in a linear-array probe; and (IV) no clear definitions of benign, malignant, and borderline lesions in the pathological diagnosis of PTs were available. According to the inclusion and exclusion criteria, 40 cases of PTs and 290 cases of FAs from female patients were finally enrolled in our study. The data came from the ultrasonic images of participants, and features were taken from those images and output as tabular data by AI.

Instruments

The ultrasound instruments used in this study comprised an Aplio 500 system (Canon Medical Systems Corp., Tochigi, Japan) with a 14L5 transducer, a Logic E9 system (GE Healthcare, Chicago, IL, USA) with a L12-5 transducer, an IU22 scanner (Philips, Amsterdam, Netherlands) with a L12-5 transducer, and a Siemens S3000 (Siemens, Munich, Germany) system with a 9L4 transducer.

The AI system used in our study was the Breast Ultrasound Intelligent Diagnosis System from the Harbin Institute of Technology. The system diagnosed tumors in

breast ultrasonic images using automatic segmentation and identification. The diagnostic sensitivity and specificity were both >90%. All lesions in the ultrasonic images were manually segmented by Sihua Niu and Xue Wang from the Peking University People's Hospital, both of whom have been engaged in the diagnosis of breast diseases diagnosis for more than 10 years. The manually segmented region was studied using grey-level gradient co-occurrence matrix analysis to acquire the morphology and textural features. Statistical analysis was then conducted among these features. An overview of the methods and procedures is provided in *Figure 1*.

Quantitative analysis

A total of 8 morphology features were enrolled in our study, which comprised circularity, height-to-width ratio, margin spicules, margin coarseness (MC), margin indistinctness, margin lobulation (ML), internal calcification, and angle between the long axis of the lesion and skin. The features were specified as follows (15):

Circularity (Cir)
Cir represented the state of lesions being similar to circles, and its calculation is shown in Eq. [1]:

$$Cir = \frac{C^2}{S} \qquad [1]$$

C was defined as the pixel number on the boundary of a lesion, which represented the perimeter of the lesion. S was the pixel number within the tumor, and it was considered as the lesion area. When the size of a perimeter was determined, the area of a circle in the plane geometry was the largest, so the closer the tumor was to a circle, the smaller the Cir was and the more regular was the tumor.

Height-to-width ratio
First, we calculated the circumscribed rectangle of the lesion and then obtained its height and width. The height (H)-to-width (W) ratio (HWR) was calculated as depicted in Eq. [2]:

$$HWR = \frac{H}{W} \qquad [2]$$

Margin spicules (MS)
MS are linearly hypoechoic structures protruding from a lesion to the periphery that are common malignant signs

of breast cancer. First, the coordinates of the margin pixels (CMPs) were converted to polar coordinates (PCs) according to centroid coordinates (CCs). The CCs referred to the position of a point relative to the centroid in a geometric structure. The PCs were determined by 1 pole and 1 polar axis. The centroid of a lesion was considered as the pole of the PC. Next, the coordinates were rearranged clockwise or anticlockwise, with the aim of all procedures being performed in the same direction. Fourier transformation was then performed to decompose the signal into the frequency spectrum. The calculation of MS is displayed in Eq. [3]:

$$MS = \frac{\sum_{\omega=0}^{\frac{\pi}{4}} R(\omega)}{\sum_{\omega=\frac{\pi}{4}}^{\pi} R(\omega)} \quad [3]$$

MC

Margin referred to the border between the lesion and surrounding tissue. MC reflected the irregularity grade of the lesion margin, which is shown in Eq. [4]:

$$MC = \frac{1}{N} \sum_{i=1}^{N} |d_i - d_{i+1}| \quad [4]$$

Here, N refers to the number of all pixels on the margin, and d_i is the distance of the i_{th} pixel on the edge to the tumor centroid in CCs. We arranged and calculated d_i on the basis of the clockwise (or anticlockwise) direction of the corresponding pixels on the margin. Both clockwise and anticlockwise could be used, and all procedures were ideally performed in the same direction.

Margin indistinctness

In the original greyscale ultrasound image, the coarse boundary of the tumor was analyzed using manually contoured images. The area around the lesion was considered as the margin region. The Sobel operator was used for calculating the pixel gradients of the margin region. The calculation of margin indistinctness (MI) is described in Eq. [5]:

$$MI = \sum_{i=1}^{n} \sum_{j=1}^{m} \sqrt{d_x(i,j)^2 + d_y(i,j)^2} \quad [5]$$

m and n represent the lesion sizes, and d_x and d_y represent the pixel gradients on the tumor margin in the horizontal and vertical directions, respectively.

ML

The PCs were constructed taking the lesion centroid as the pole. We converted the CMPs to PCs, and created a PC sequence along the clockwise (or anticlockwise) direction. Both clockwise and anticlockwise direction could be used, and all procedures were ideally performed in the same direction. In order to reduce the influence of image noise, the median filter of frame size 21 was employed, and a polynomial of 20 degrees was used to fit the sequence. A frame size of 21 and a polynomial of 20 degrees were determined by experiments as the best choices to balance the accuracy of margin curve fitting and the complexity of calculation. By reducing the influence of image noise, we were able to obtain accurate ML. The total of the maximum and minimum values was regarded as the result of ML shown in Eq. [6]:

$$ML = sum(f_{max}) + sum(f_{min}) \quad [6]$$

Internal calcification

The region surrounding a lesion was set to 0 pixels, and the grey level of the lesion was binarized according to the average grey level. Internal calcifications (IC) lay in regions with the maximum grey level. Then, the interference pixels were removed from the binarized image. Connected white regions in the binarized image were considered as IC of the lesion. For example, if there were 3 connected regions of white spots, the IC of the lesion was 3.

Angle between the long axis of the lesion and skin (ALS)

The ellipse-fitting algorithm was used to fit the lesion margin to an ellipse. The intersection of two symmetrical axes was considered the ellipse center. The long axis, short axis, fitted ellipse center, and angle between the long axis and horizontal (ALS, θ) were definite. The transformations were conducted as shown in Eq. [7]:

$$\begin{cases} \theta, & 0 \leq \theta \leq \frac{\pi}{2} \\ \pi - \theta, & \frac{\pi}{2} \leq \theta \leq \pi \\ \theta - \pi, & \pi \leq \theta \leq \frac{3\pi}{2} \\ 2\pi - \theta, & \frac{3\pi}{2} \leq \theta \leq 2\pi \end{cases} \quad [7]$$

A total of 3 textural features were enrolled in our research, which comprised energy, grey entropy, and grey mean. First, we converted greyscale images to gradient images. L_S was the maximum grey value of the grey image. L_g was the maximum gradient value of the gradient image. $H(i, j)$ was defined as pixels with a grey level of i and a gradient level of j simultaneously. $H(i, j)$ was then

2056

Niu et al. Difference between phyllodes tumors and fibroadenomas

Table 1 Cases of each kind of ultrasound instruments.

Tumors	Logio E9	Philips IU22	Aplo 500	Siemens S3000	Total
PTs	30	9	0	1	40
FAs	265	6	18	1	290
Total	295	15	18	2	330

PTs, phyllodes tumours; FAs, fibroadenomas.

normalized and converted to $P(i, j)$, and $P(i, j)$ was used to calculate energy, grey entropy, and grey mean, which are shown in Eqs. [8], [9], and [10], respectively (15):

Energy (E)

E was the square sum of the elements in the region of interest. It reflected the coarseness of the texture in the lesions. E was calculated according to Eq. [8]:

$$E = \sum_{i=1}^{L_s}\sum_{j=1}^{L_g} P(i, j)^2 \qquad [8]$$

Grey entropy (GE)

GE measured the amount of information in an image. The more scattering there was, the less information was detected, and the smaller was the GE. The GE was calculated using Eq. [9]:

$$GE = -\sum_{i=1}^{L_s}\left(\sum_{j=1}^{L_g} P(i, j)*log\sum_{j=1}^{L_g} P(i, j)\right) \qquad [9]$$

Grey mean (GM)

GM described the average grey level of an image. The larger the GM, the brighter the whole image was. The GM was calculated according to Eq. [10]:

$$GM = \sum_{i=1}^{L_s} i*\sum_{j=1}^{L_g} P(i, j) \qquad [10]$$

Statistical analysis

The statistical software SPSS version 17.0 software (IBM Corp., Armonk, NY, USA) was used for data analyses. The normality test showed that the data of morphology and textural features were abnormally distributed. The median, 25th percentile, and 75th percentile were used to describe the data. A nonparametric test (Mann-Whitney) was employed to compare all morphology and textural features between PTs and FAs and determine the features with statistical significance. Our data were dichotomous,

so receiver operating characteristic (ROC) curve analysis was used to calculate the area under curve (AUC), which could evaluate the diagnostic performance of the features. According to the biggest Youden index, we determined the cutoff values on ROC curves for the differential diagnosis of PTs and FAs and recorded their sensitivity, specificity, positive predictive value (PPV), and NPV. Sensitivity reflected the rate of PT diagnosis among the true PTs, and specificity reflected the rate of not diagnosing PTs among those which were not PTs using the cutoff value. The PPV represented the rate of true PTs for tumors diagnosed as PTs by features. The NPV represented the rate of diagnosis as not true PTs for tumors also diagnosed as not PTs by features. Cutoff values were important, and they supplied quantitative diagnosis information. A P value <0.05 was considered to indicate a statistically significant difference.

Results

A total of 40 cases were PTs, with a median age of 39 years [28–52], and a median size of 3.1 cm [1.5–4.8]. Among the 40 PTs, 26 were benign PTs, 12 were borderline PTs, and 2 cases were malignant PTs. Some 290 participants had FAs, with a median age of 38 years [21–53], and a median size of 1.4 cm [0.6–3.8].

We used 4 kinds of ultrasound instruments in our research. Cases of each kind of ultrasound instrument are shown in *Table 1*. Most cases (295/330, 89.4%) were scanned with the Logio E9 instrument.

Comparison of all morphology and textural features between PTs and FAs

Data of the morphology and textural features of PTs and FAs are shown in *Table 2*. Statistically significant differences (P<0.05) were found in the HWR, ML, E, and GE between PTs and FAs. The PTs had a larger HWR and E, more MLs, and smaller GE than FAs (*Figures 2,3*). No statistically significant differences existed in Cir, MS, MC, MI, GM, IC,

Table 2 Comparison of morphology and textural features between PTs and FAs

Features	PTs			FAs			P
	Median	25th percentile	75th percentile	Median	25th percentile	75th percentile	
Cir	18.805	16.945	21.250	20.015	17.583	22.880	0.109
HWR	0.635	0.553	0.708	0.520	0.430	0.623	*<0.001*
MS	18.525	11.815	27.043	18.780	13.168	25.835	0.818
MC (pixel)	1.340	1.063	1.803	1.490	1.198	1.940	0.082
MI (E6)	2.019	1.433	3.011	1.854	1.288	2.429	0.122
ML	7.000	5.500	9.000	7.000	5.000	9.000	*0.021*
E	0.050	0.030	0.080	0.030	0.020	0.040	*<0.001*
GE	2.120	1.993	2.238	2.280	2.160	2.363	*<0.001*
GM	0.465	0.350	0.750	0.485	0.388	0.630	0.731
IC	1.500	0.000	5.750	1.000	0.000	3.000	0.423
ALS (degree)	7.050	1.710	14.063	5.605	2.120	11.385	0.701

The results are presented as medians, 25th percentiles, and 75th percentiles. The P values are for the nonparametric test results of the morphology and textural features of 40 PTs and 290 FAs. Values in italic denote statistically significant features. PTs, phyllodes tumors; FAs, fibroadenomas; Cir, circularity; HWR, height-to-width ratio; MS, margin spicules; MC, margin coarseness; MI, margin indistinctness; ML, margin lobulation; E, energy; GE, grey entropy; GM, grey mean; IC, internal calcification; ALS, angle between the long axis of the lesion and skin.

and ALS between PTs and FAs.

Evaluation of the diagnostic performance of morphology and textural features in differentiating PTs and FAs

Data of the diagnostic performance of single features for the differentiation between PTs and FAs are shown in *Table 3* and *Figure 4*. The ROC curve analysis of single features showed that the AUC of GE (0.759) was the largest of the four features with a statistical difference. When the cutoff value was 2.295, the sensitivity, specificity, PPV, and NPV of GE were 0.925, 0.459, 0.978, and 0.190, respectively. To balance the complexity of models and improve the AUC due to increasing the features and the results of single feature, we obtained the optimal 3 combinations among the features, which are shown in *Table 4* and *Figure 4*. The AUC of the combination of HWR, MI, ML, E, GE, and IC was 0.868, and the sensitivity, specificity, PPV, and NPV were 0.734, 0.900, 0.982, and 0.316, respectively.

Discussion

Our research showed that there was a statistically significant difference in HWR (P<0.001), which was likely due to the rapid growth of PTs (2,16). The HWR of PTs was relatively larger than that of FAs. The specific value of the HWR

could be devised using AI as opposed to simply stating whether the HWR was greater or less than 1. From this perspective, AI exceeds the performance of human eyes.

In our study, quantitative analysis of AI indicated that the ML of PTs was different from that of FAs. The cause for this may be that PTs showed more leaf-like architectures and much faster growth (12). The ML was a main ultrasonic feature of PTs (17). Our findings were consistent with previous reports (12,17).

The E and GE reflect the heterogeneity of echoes in the lesions. The E is the square sum of the elements of the grey-level gradient co-occurrence matrix. The greater the E, the more heterogeneous the echo. The GE reflects the complexity and nonuniformity of image texture. A greater GE means more information in an image, which is more homogeneous (18). In our study, PTs showed a larger E and a smaller GE than did FAs. Compared with FAs, PTs are more likely to present with cystic areas (12,17), and the internal echo of PTs is more heterogeneous. If the internal components are more complex, the scatter will increase, causing backscatter variation and reducing GE. It has been reported that the presence of a cystic area indicates that the risk of PTs is 16.5 times higher than that of FAs (4). In one report, the cystic area was used to differentiate PTs
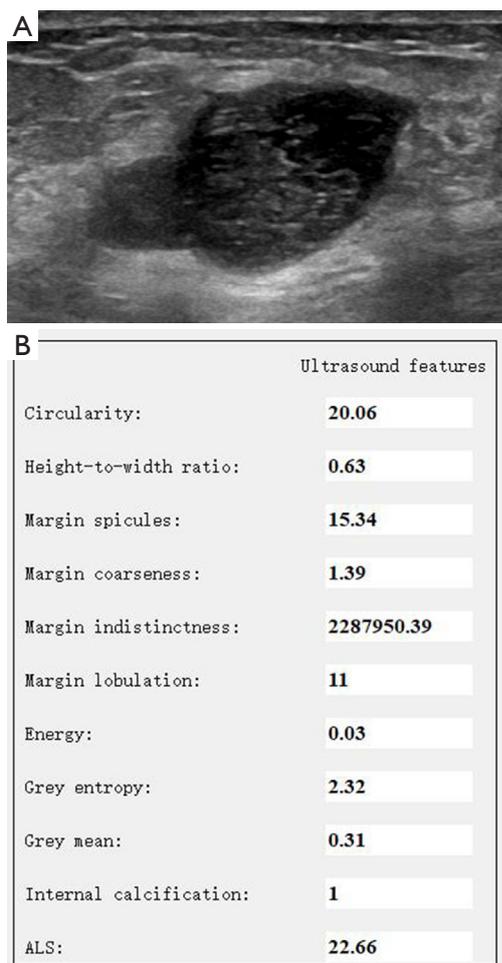
Figure 2 A phyllodes tumor in a 41-year-old woman. The original ultrasonic image of the lesion is shown in (A). The data of morphology and textural features of the lesion are shown in (B).
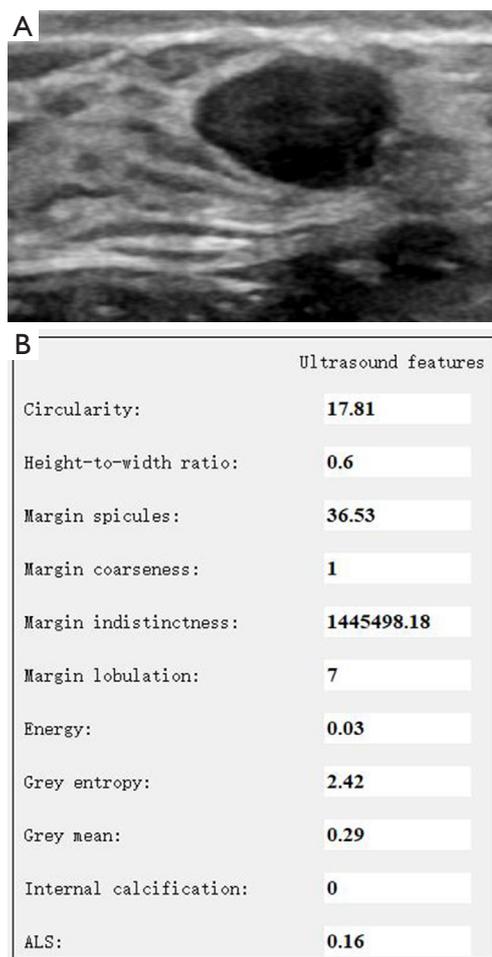
Figure 3 A fibroadenoma in a 32-year-old woman. The original ultrasonic image of the lesion is shown in (A). The data of morphology and textural features of the lesion are shown in (B).

Table 3 Diagnostic performance analysis of the four statistically significant features for the differentiation between PTs and FAs

Features	AUC	Cutoff value	Sensitivity	Specificity	PPV	NPV
HWR	0.733	0.525	0.875	0.534	0.969	0.206
ML	0.610	6.000	0.750	0.445	0.928	0.157
E	0.758	0.035	0.725	0.662	0.946	0.228
GE	0.759	2.295	0.925	0.459	0.978	0.190

PTs, phyllodes tumors; FAs, fibroadenomas; HWR, height-to-width ratio; ML, margin lobulation; E, energy; GE, grey entropy; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

from FAs, yielding a sensitivity of 21.8% and a specificity of 98.3% (4). Therefore, the more heterogeneous the lesion, the more likely it is to be a PT, particularly when a cystic area is present.

We aimed to provide an exploratory reference for the differential diagnosis of PTs and FAs using AI. In this study, when using ROC curve analysis of the HWR, ML, E, and GE, we found that the AUC of GE was the largest.

At a cutoff value of 2.295, the sensitivity and PPV of GE for the diagnosis of PTs were high but the specificity and NPV were very low. This result might have been caused by the great overlap of GE values between PTs and FAs. It is obviously difficult to diagnose a disease based on a solitary feature. Our results showed that the optimal AUC existed in combination A (HWR, MI, ML, E, GE and IC). Combination A improved the specificity, PPV, and NPV, but the NPV was still low. This was also potentially due to the overlap of features between the two kinds of tumors. Using GE and combination A to diagnose PTs, we could detect the maximum number of PTs to reduce the omission diagnosis of PTs, allowing active treatment to be initiated to reduce recurrence and metastasis of PTs. Meanwhile, the specificity and NPV were low for non-PT lesions. In this case, the low

specificity and NPV were acceptable, because many FAs required surgery to thoroughly remove the lesions.

Comparison with other studies

There are many similarities between PT and FA tumors. The pathological difference between them usually lies in the more obvious intraductal growth, leaf-like architectures, and abundant cellular stroma in PTs (19). However, challenges remain in differentiating PTs from FAs by imaging (2,8,14,16,20,21).

AI extracts meaningful information from ultrasonic images by quantitative analysis of morphology and textural features of lesions, which can not only be used to diagnose breast diseases but also contribute to the categorization of tumor pathological types (22-24). Presently, most AI studies focus on the diagnosis of benign and malignant breast lesions (11,25,26); only a few papers have focused on the differential diagnosis of PTs and FAs (12-14). Results from one study showed that the ML and homogeneity of the PTs were different from those of the FAs, and the combination of clinical features, MRI features, and textural features could improve the sensitivity and specificity of the differential diagnosis between PTs and FAs (12). Another study has shown that the deep learning of AI could improve the accuracy of distinguishing PTs from FAs (14). Textural analysis of ultrasonic images was a potential diagnostic tool for the differential diagnosis of benign PTs, complex FAs, and simple FAs, and differences existed in E and GE among the three types (13,14). Our results are consistent with the above-mentioned findings. Although we chose some small PTs, ML and the homogeneity of the lesions were still the points of distinction between PTs and FAs. Regardless of lesion size, the biological behavior of the PT determines its ultrasonic characteristics.

Strengths and limitations of this study

We attempted to use AI to analyze very similar tumors
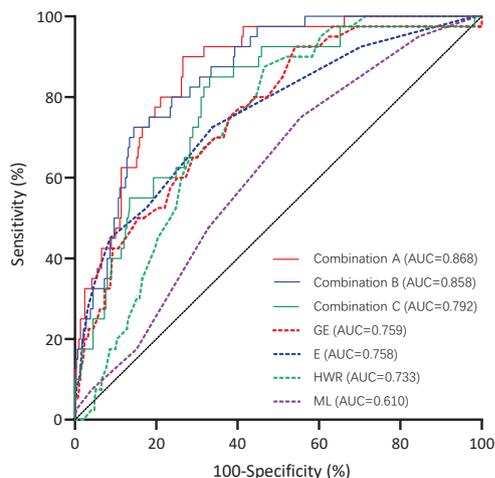
Figure 4 The ROC curve analysis from statistically significant features and combined features for the differentiation between phyllodes tumors and fibroadenomas. Combination A: HWR, MI, ML, E, GE, IC. Combination B: HWR, MI, GE, IC. Combination C: HWR, GE. ROC, receiver operating characteristic; HWR, height-to-width ratio; MI, margin indistinctness; ML, margin lobulation; E, energy; GE, grey entropy; IC, internal calcification.

Table 4 Diagnostic performance of combined features for the differentiation between PTs and FAs

Item	Combined features	AUC	Cutoff value	Sensitivity	Specificity	PPV	NPV
Combination A	HWR, MI, ML, E, GE, IC	0.868	0.906	0.734	0.900	0.982	0.316
Combination B	HWR, MI, GE, IC	0.858	0.851	0.855	0.725	0.958	0.408
Combination C	HWR, GE	0.792	0.907	0.669	0.850	0.970	0.262

PTs, phyllodes tumors; FAs, fibroadenomas; HWR, height-to-width ratio; MI, margin indistinctness; ML, margin lobulation; E, energy; GE, grey entropy; IC, internal calcification; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

2060

Niu et al. Difference between phyllodes tumors and fibroadenomas

and provide quantitative information for their differential diagnosis. The results showed that AI could be used for the differential diagnosis of PTs and FAs.

Our study had some limitations. Firstly, we selected PTs <5 cm, and the morphology and textural features of larger PTs remained unexplored. In the future, we will compare the textural and morphological features of PTs of different sizes. Secondly, the sample size of 40 PTs in this study was relatively small, especially for the analysis of diagnostic efficiency. We will continue to collect PTs and investigate the differential diagnosis between PTs and FAs. Thirdly, due to the error of automatic segmentation, we explored the differential diagnosis between PTs and FAs using manually contoured images, and the diagnostic efficiency based on automatic recognition of the AI diagnosis system was not evaluated. Fourthly, on account of the small number of each pathological type of PT, we did not compare different pathological types of PTs with FAs. We will continue to collect different pathological types of PT, and in the future, these aspects will be assessed.

Conclusions

Quantitative analysis of AI can discern the subtle differences in morphology and textural features between PTs and FAs. Comprehensive consideration of multiple features contributes to the diagnosis of PTs and FAs. With the development of AI, by training with a large sample and an improved calculation method we may be able to supply more accurate information to diagnose PTs and FAs.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/qims-20-919). All authors report receiving grants from the Beijing Municipal Science & Technology Commission during the conduct of this study.

Ethical Statement: This study was approved by the Ethics Board of the Peking University People's Hospital (No. 2018PHB179-01), and informed consent was given by all participants.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Rowell MD, Perry RR, Hsiu JG, Barranco SC. Phyllodes tumors. Am J Surg 1993;165:376-9.
2. Tan H, Zhang S, Liu H, Peng W, Li R, Gu Y, Wang X, Mao J, Shen X. Imaging findings in phyllodes tumors of the breast. Eur J Radiol 2012;81:e62-9.
3. Kalambo M, Adrada BE, Adeyefa MM, Krishnamurthy S, Hess K, Carkaci S, Whitman GJ. Phyllodes Tumor of the Breast: Ultrasound-Pathology Correlation. AJR Am J Roentgenol 2018;210:W173-W179.
4. Xiao M, Zhu Q, Jiang Y, Li J, Wang H, Zhang J, You S, Liu H. Local Recurrent Phyllodes Tumors of the Breast: Clinical and Sonographic Features. J Ultrasound Med 2015;34:1631-8.
5. Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, Van der Vijver MJ. WHO classification of tumours of the breast. 4th ed. Lyon: IARC, 2012.
6. Suvannarerg V, Chitchumnong P, Apiwat W, Lertdamrongdej L, Tretipwanit N, Pisarnturakit P, Sitthinamsuwan P, Thiravit S, Muangsomboon K, Korpraphong P. Diagnostic performance of qualitative and quantitative shear wave elastography in differentiating malignant from benign breast masses, and association with the histological prognostic factors. Quant Imaging Med Surg 2019;9:386-98.
7. Pediconi F, Galati F. Breast cancer screening programs: does one risk fit all? Quant Imaging Med Surg 2020;10:886-90.
8. Mishra SP, Tiwary SK, Mishra M, Khanna AK. Phyllodes tumor of breast: a review article. ISRN Surg 2013;2013:361469.
9. Park HL, Kwon SH, Chang SY, Huh JY, Kim JY, Shim JY, Lee YH. Long-term follow-up result of benign phyllodes tumor of the breast diagnosed and excised by ultrasound-guided vacuum-assisted breast biopsy. J Breast Cancer

2012;15:224-9.

10. Shan J, Alam SK, Garra B, Zhang Y, Ahmed T. Computer-Aided Diagnosis for Breast Ultrasound Using Computerized BI-RADS Features and Machine Learning Methods. Ultrasound Med Biol 2016;42:980-8.

11. Moon WK, Lo CM, Chang JM, Huang CS, Chen JH, Chang RF. Quantitative ultrasound analysis for classification of BI-RADS category 3 breast masses. J Digit Imaging 2013;26:1091-8.

12. Mai H, Mao Y, Dong T, Tan Y, Huang X, Wu S, Huang S, Zhong X, Qiu Y, Luo L, Jiang K. The Utility of Texture Analysis Based on Breast Magnetic Resonance Imaging in Differentiating Phyllodes Tumors From Fibroadenomas. Front Oncol 2019;9:1021.

13. Basara Akin I, Ozgul H, Simsek K, Altay C, Secil M, Balci P. Texture Analysis of Ultrasound Images to Differentiate Simple Fibroadenomas From Complex Fibroadenomas and Benign Phyllodes Tumors. J Ultrasound Med 2020;39:1993-2003.

14. Stoffel E, Becker AS, Wurnig MC, Marcon M, Ghafoor S, Berger N, Boss A. Distinction between phyllodes tumor and fibroadenoma in breast ultrasound using deep learning image analysis. Eur J Radiol Open 2018;5:165-70.

15. Niu S, Huang J, Li J, Liu X, Wang D, Zhang R, Wang Y, Shen H, Qi M, Xiao Y, Guan M, Liu H, Li D, Liu F, Wang X, Xiong Y, Gao S, Wang X, Zhu J. Application of ultrasound artificial intelligence in the differential diagnosis between benign and malignant breast lesions of BI-RADS 4A. BMC Cancer 2020;20:959.

16. Bozkurt H, Karakaya IB, Aktas E, Irkorucu O. Coexistence of phylloides tumour and invasive ductal cancer in the breast. Niger J Clin Pract 2019;22:1169-71.

17. Chao TC, Lo YF, Chen SC, Chen MF. Phyllodes tumors of the breast. Eur Radiol 2003;13:88-93.

18. Tsui PH, Chen CK, Kuo WH, Chang KJ, Fang J, Ma HY, Chou D. Small-window parametric imaging based on information entropy for ultrasound tissue characterization. Sci Rep 2017;7:41004.

19. Jacklin RK, Ridgway PF, Ziprin P, Healy V, Hadjiminas D, Darzi A. Optimising preoperative diagnosis in phyllodes tumour of the breast. J Clin Pathol 2006;59:454-9.

20. Li LJ, Zeng H, Ou B, Luo BM, Xiao XY, Zhong WJ, Zhao XB, Zhao ZZ, Yang HY, Zhi H. Ultrasonic elastography features of phyllodes tumors of the breast: a clinical research. PLoS One 2014;9:e85257.

21. Kim GR, Kim EK, Yoon JH, Kim MJ, Moon HJ. Recurrence Rates of Benign Phyllodes Tumors After Surgical Excision and Ultrasonography-Guided Vacuum-Assisted Excision. Ultrasound Q 2016;32:151-6.

22. Wu T, Sultan LR, Tian J, Cary TW, Sehgal CM. Machine learning for diagnostic ultrasound of triple-negative breast cancer. Breast Cancer Res Treat 2019;173:365-73.

23. Klimonda Z, Karwat P, Dobruch-Sobczak K, Piotrzkowska-Wróblewska H, Litniewski J. Breast-lesions characterization using Quantitative Ultrasound features of peritumoral tissue. Sci Rep 2019;9:7963.

24. Zhang L, Li J, Xiao Y, Cui H, Du G, Wang Y, Li Z, Wu T, Li X, Tian J. Identifying ultrasound and clinical features of breast cancer molecular subtypes by ensemble decision. Sci Rep 2015;5:11085.

25. Venkatesh SS, Levenback BJ, Sultan LR, Bouzghar G, Sehgal CM. Going beyond a First Reader: A Machine Learning Methodology for Optimizing Cost and Performance in Breast Ultrasound Diagnosis. Ultrasound Med Biol 2015;41:3148-62.

26. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. Clin Radiol 2019;74:357-66.