

RESEARCH

Open Access



Discrimination between phyllodes tumor and fibro-adenoma: Does artificial intelligence-aided mammograms have an impact?

Sahar Mansour^{1*} , Rasha Kamel¹, Ahmed Marey², Christiane Hunold² and Ahmed Yousry³

Abstract

Background: The indulgence of artificial intelligence (AI) has been considered recently in the work up for the detection and diagnosis of breast cancer through algorithms that could supply diagnosis as the radiologist do. The algorithm learns from a supervised and continuous input of large and new data sets unlike the standard programming, which requires clear step-by-step instructions. The aim of this study is to assess the ability of AI scanned mammograms to aid the ultrasound in the discrimination between phyllodes tumors and fibro-adenomas.

Results: This is a retrospective analysis included 374 proven phyllodes tumors (PT) and fibro-adenomas (FA). Digital mammogram and breast ultrasound was performed for all the cases and each breast was given a "Breast Imaging Reporting and Data System" (BI-RADS) score. Included mammograms were scanned by AI with resultant a qualitative heatmap and a quantitative abnormality scoring of suspicion percentage.

The study included 164 PT (43.9%) and 210 FA (56.1%). BI-RADS category 2 was assigned in 40.1%, category 3 in 38.2%, category 4 in 18.5% and category 5 in 3.2% with median value of the AI abnormality scoring of 23%, 44%, 65% and 90% respectively. Sensitivity and specificity of the conventional imaging were 59.2% and 75.8% respectively. The AI abnormality scoring of 49.5% upgraded the sensitivity to 89.6% and specificity to 94.8% in the ability to discriminate PT from FA masses.

Conclusion: Artificial intelligence-aided mammograms could be used as method of distinction between PT from FA detected on sono-mammogram. The color hue and the quantification of the abnormality scoring percentage could be used as a one setting method for specification and so guide clinicians in their decision of conservative management or the choice of the surgical procedure.

Keywords: Artificial intelligence, Colour hue, Abnormality scoring, Phylloids, Fibroadenoma, Digital mammogram

Background

It has been reported that phyllodes tumors are usually benign and consist of the stromal component and the epithelial component that may progress to malignancy [1].

There are histopathological similarities between fibro-adenomas and benign and malignant phyllodes tumors, moreover there are no pathognomonic radiological finding for distinction as well [2].

Fibro-adenomas can be followed without the need for further work up or be subjected to simple removal, but phylloids tumors must be surgically removed with wide

*Correspondence: sahar_mnsr@yahoo.com

¹ Women's Imaging Unit, Radiology Department, Kasr ElAiny Hospital, Cairo University, ElManial, Cairo, Egypt

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

adequate margins to avoid local recurrence. The rate of local recurrence of the benign and borderline subtypes of the phyllodes tumors are 20 and more than 25%, respectively and approximately 25% of the malignant phyllodes tumors metastasize [3].

Recently, the advances in deep learning presented a valuable participation to upgrade the performance of digital mammogram for breast imaging [4].

The goal from introducing technology in medicine was to make algorithms able to make decisions like physicians [5].

The role of artificial intelligence (AI) in the work up for the detection of breast cancer is becoming clear [6].

Artificial intelligence is to make computers/computer programs smart enough to imitate the human mind behavior. In machine learning, the continuous exposure to new data improves and modulates the algorithm, there is no need for explicit programming [7].

The article discusses the feasibility to use artificial intelligence (AI) as a supportive tool that could be used to discriminate between phyllodes tumor or fibro-adenoma clearly detected at the conventional imaging (mammogram aided by ultrasound) and so could guide the physicians in their choice for the proper surgical procedure.

Methods

Patients

The current work evaluated a total of 374 masses of 350 females (single breast was involved in 326 patients and the remaining 24 patients presented with bilateral breast affection) over the period of August 2019 till July 2021. Included mammograms was supported by ultrasound assessment and presented masses proved to be phyllodes tumors (PT) in 164 and fibro-adenomas (FA) in 210 lesions. Masses were palpable in 332 (88.8%) and non-palpable in the remaining 42 masses. Mammograms included in the study presented one mass at a time.

The study is a retrospective analysis, ethics committee approved and received waiver of the informed consent from the included patients. The age of the included patients ranged between 30 and 54 years old with mean age 42.2 ± 6.6 SD.

Inclusion criteria

Solid (purely solid or complex) breast masses proved by a tissue core biopsy of 14 G needle and surgery to be either phyllodes tumor or fibro-adenomas clearly detected by the combined evaluation of mammogram and ultrasound.

Patients with benign or probably benign looking masses underwent biopsy and surgery at the request of the patient, for cosmetic reasons, or because of a

recommendation of the referring physician to guide the type of excision.

Exclusion criteria

Masses that were shown only on the ultrasound and not obvious on the mammograms.

Masses with no available abnormality scoring percentage on mammograms.

Diagnosed tumors without pathological documentation.

No clear definitions of benign and malignant phyllodes in the pathology report.

Methods

All patients were imaged by digital mammography machine (Amulet Innovality, Fujifilm Global Company, Japan) and the traditional cranio-caudal and medio-lateral oblique views were taken per breast.

Since some masses showed obscured borders or were not obviously seen on the mammogram, ultrasound examinations (the used machine was HS60 Samsung ultrasound, Korea, 2019) were performed to support (confirm or upgrade) the “Breast Imaging Reading and Data System” (BI-RADS) category for the detected masses.

Artificial intelligence software scanned the included mammograms (Lunit Insight, FDA approved, and version 2019) which were provided with an established algorithm to read the mammogram images.

Image analysis

Mammograms were interpreted by three radiologists in individual sessions that were blinded about the pathology of the included masses. Two primary evaluators (each had 20 years experience in breast imaging) assessed the mammograms and performed the complementary ultrasound. Afterwards, a third reader (35 years experience) re-interpreted the mammograms to settle the decision in case of debate or confirm diagnosis in case of matched opinions.

Morphology descriptors of the included masses as elicited by the conventional imaging (digital mammogram and complementary ultrasound) were size, margin (circumscribed, lobulated, indistinct), shape (rounded/oval, irregular), internal structure (purely solid, complex with few cystic spaces, calcifications).

Following the ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System [8], the included masses were assigned BI-RADS categories 2, 3, 4 or 5 that were given with regard the combined analysis of the digital mammogram and the hand-held ultrasound.

In view of Plaza et al. [9], breast masses that displayed one or more of these features: large size of >4 cm,

lobulated margin, irregular shape, heterogeneous texture, presence of cystic spaces and absence of microcalcifications were characteristics of PA rather than of FA.

Mammograms were rescanned by the AI software where an algorithm was developed based on deep convolutional neural networks (CNNs). ResNet-34, one of the most popular CNN architectures, was used as a backbone network. The algorithm training consists of two stages: patch-level training from scratch for learning low-level features, followed by image-level fine tuning from the stage-1 model for learning high-level context (stage 2).

Batch-instance normalization was performed to overcome variance pixel-level characteristics and a deconvolution module was additionally adopted to decrease the rate of the false positives.

For an input mammogram image (i.e., one of the four views) the AI algorithm provided:

- (I) A qualitative parameter; pixel-level abnormality scores as a “heatmap” where a color hue overlaid the masses (i.e., abnormalities) detected on the mammogram images. The colors ranged between cold intensities as blue or green for lesions with low suspicion and high intensity colors as yellow, orange, and red for masses with high suspicion of activity.
- (II) Abnormalities detected by the AI software on the mammogram were assigned a numerical value which represents a degree of confidence for the suspicion. The abnormality scores are floating-point values between 0 and 1 which is estimated by an auto-applied percentages that ranged from 1 to 100% (the number 100% represented the upper limit value of suspicion).

The used AI software (Lunit INSIGHT MMG) provides four-view heatmaps and an abnormality score per breast (i.e., the maximum of the craniocaudal and mediolateral oblique abnormality scores) for each input mammogram.

The performance of the AI algorithm was supervised and maintained by a specialist of AI informatics (10 years experience).

Statistical analysis

Data were coded and entered using the statistical package SPSS (Statistical Package for the Social Sciences) version 26 (IBM Corp., Armonk, NY, USA). Data was summarized using mean, standard deviation, minimum and maximum in quantitative data and using frequency (count) and relative frequency (percentage) for categorical data. Standard diagnostic indices including sensitivity, specificity, positive predictive value (PPV),

negative predictive value (NPV) and diagnostic efficacy were calculated.

The statistical dispersion was measured by the inter-quartile range to enhance the accuracy of the dataset statistics by dropping lower contribution, outlying points. For comparing categorical data, Chi square test was performed, and *P* value was obtained, where values less than 0.05 were considered as statistically significant. The inter-observer variability was measured (to calculate measurement error intrinsic to between-observer difference) using kappa indices.

The cut-off value for the abnormality scoring percentage was detected with the aid of the “receiver operating characteristic” (ROC) curve. Confidence interval percentage (CI %) was done for the range of the abnormality scoring values elicited by the AI software where the narrower the interval (upper and lower values), the more precise is the AI estimate.

Results

The included mammograms (presented 374 masses) of 350 females showed breast density of ACR a (predominantly fatty) in 24% ($n=84/350$), ACR b (i.e. scattered fibro-glandular tissue) in 27.2% ($n=95$), ACR c (i.e. heterogeneously dense) in 39.1% ($n=137/350$), and ACR d (extremely dense) in 9.7% ($n=34/350$). Right breast was involved in 52.7% ($n=197$) and left breast in 47.3% ($n=177$).

The mean size of the proven PT and FA masses at surgery was 3.55 cm (min. 0.60 cm, max. 10.00 cm) and was 2.94 cm (min. 0.70 cm, max. 11.3 cm) on the ultrasound.

Since the resultant AI image is an overlay scan of the mammogram, so actual estimation of the lesion size was not applicable by the AI system that is why the lesion size was only presented with regard the conventional imaging and was not considered in the correlation between the used breast imaging tools and the AI- scanned mammograms.

PT compromised 43.9%, while FA presented 56.1% of the studied masses. Included phylloids masses were found benign in 152 masses and the remaining twelve masses were malignant tumors ($n=12/374$; 3.2%).

Suspicious morphology descriptors presented the near frequency in case of the PT and FA masses. These descriptors included: the irregular margin and the indistinct shape. Irregular margin was more common in the PA (34.7%) compared to the FA (29.5%), while indistinct shape was more frequently seen with the FA (45%) than the PT (42%).

The heterogeneous texture and the complex appearance compromised the internal features of 82.3% of the PT and only 42.8% of the FA masses.

Table 1 The various morphology descriptors that were detected at the conventional imaging of the PT and FA included in the study

Morphology descriptors	Phylloids tumours (n = 164)	Fibroadenomas (n = 210)
<i>Margin</i>		
Circumscribed	107 (75.6%)	130 (61.9%)
Lobulated	23 (14%)	50 (23.8%)
Indistinct	17 (10.4%)	30 (14.3%)
<i>Shape</i>		
Rounded/oval	100 (61%)	115 (55%)
Irregular	64 (39%)	95 (45%)
<i>Internal structure</i>		
Purely solid, homogenous	29 (17.7%)	120 (57.2%)
Purely solid, heterogeneous	94 (57.3%)	87 (41.4%)
Calcifications	18 (11%)	31 (14.8%)
Complex with cystic spaces	41 (25%)	3 (1.4%)

The various morphology descriptors of the PT and FA included in the study presented on the conventional imaging (mammogram and ultrasound) were displayed by Table 1.

Masses were diagnosed on imaging basis as likely benign in 150 masses (40.1%), probably benign in 143 masses (38.2%), probably malignant in 69 masses (18.5%) and likely malignant in 12 masses (3.2%). The median value for the AI abnormality scoring for each category was calculated.

Table 2 displayed the correlation between the BI-RADS category of the included PT and FA masses as evaluated by the mammogram and ultrasound and the AI assigned percentage of the abnormality scoring.

Reliability of the radiographic diagnosis of the included masses was better between readers when AI informatics was added to the sonomammogram in the evaluation 0.58 (range 0.52–0.66) ($\kappa=0.59$, range between 0.52 and 0.66) than when readers assessed masses only by mammogram and ultrasound ($\kappa=0.93$, range between 0.89 and 0.97).

According to the morphologic descriptors noted at the combined mammogram and ultrasound, readers diagnosed 277 FA and 97 PT masses. Resultant sensitivity and specificity of the conventional imaging to discriminate PT from FA masses were 59.2% and 75.8% respectively.

To determine a cut-off value for the abnormality scoring percentage that can suggest PT versus FA breast masses, the point on ROC curve was used. AI abnormality scoring of 49.5% yielded a sensitivity of 89.6% and specificity of 94.8% in the discrimination of the PT from FA.

Since PT is the entity of the more concern and requires wide excision surgery unlike FA that could be followed up so it was considered as the positive cases.

Table 3 shows discrimination ability of the AI abnormality scoring to define PT from FA in the studied masses.

In the current work, AI displayed 17 false negative and 11 false positive masses when correlated with the pathology outcome.

Discussion

Phylloids and fibro-adenomas are two benign breast pathologies which had a very close histology at the beginning, yet later their pattern of growth does divert. PT had more positive proliferative cellular nuclear antigen, Ki-67 and androgen receptors than found in FA. Moreover, PT

Table 2 The correlation between the AI assigned percentage of the abnormality scoring and the BI-RADS category of the included breast masses (PT and FA)

AI abnormality score *BI-RADS category	No	Median	1st Quartile	3st Quartile	Inter-quartile range	P value
Likely benign 2	150 (45 PT, 105 FA)	23	16.5	31	14.5	< 0.001
Probably benign 3	143 (50 PT, 93 FA)	44	26	70	44	0.741
Probably malignant 4	69 (57 PT, /12 FA)	81.5	71	88	17	0.020
Likely malignant 5	12 (12 PT)	90	77	95	16	< 0.001

Table 3 Discrimination ability of AI abnormality score to predict PT from FA using the pathology results as the gold reference

Area	Cut off value of abnormality scoring %	P value	Sensitivity	Specificity	Accuracy	95% Confidence Interval	
						Lower bound	Upper bound
0.937	49.5	< 0.001	89.6%	94.8%	92.5%	0.898	0.975

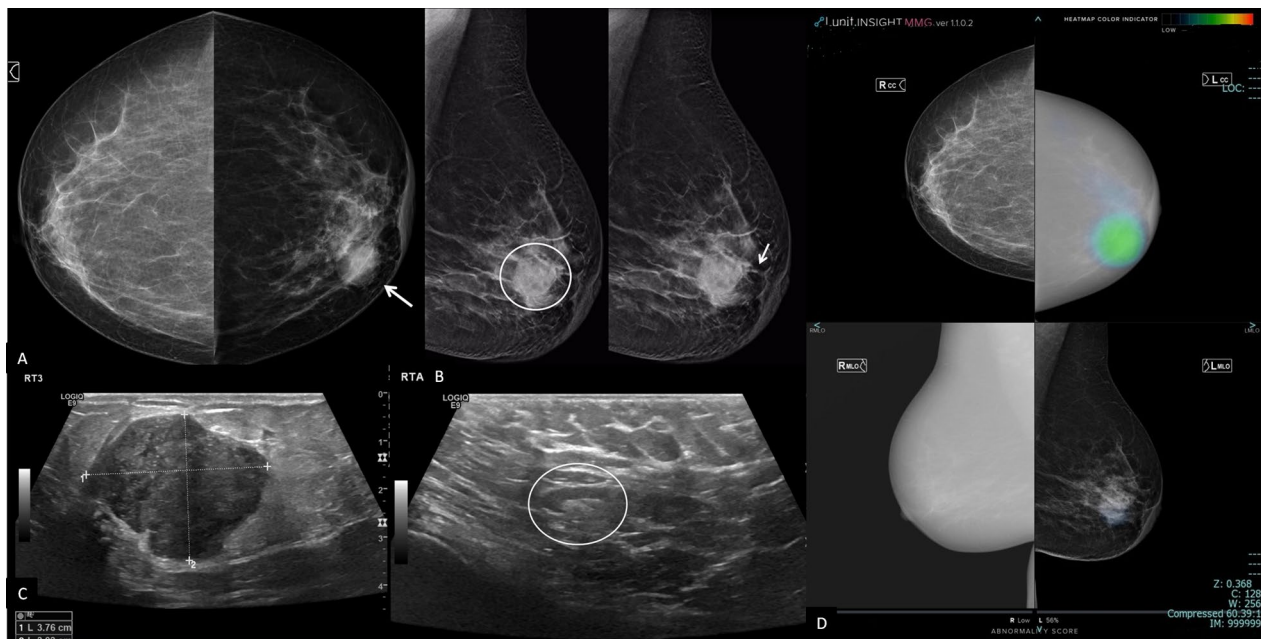


Fig. 1 Female patient 50 years old with left breast benign phylloids. **A** Bilateral digital mammogram (cranio-caudal view). There is a left breast upper inner focal rounded dense mass (arrow), that showed associate secondary signs suspicious of malignancy of increased density and coarsening of the surrounding parenchyma (suggestive peri-tumoral edema) and focal areolar dermal thickening. **B** Tomosynthesis slices of the left breast (MLO view) showed that the mass had a circumscribed margin (circle), yet there is related thickened Cooper's ligament (arrow). **C** Ultrasound images displayed large about 4 cm solid mass with irregular margin and the ipsilateral axilla displayed an average sized likely reactionary lymph node (white circle). **D** AI scanned mammogram image marked the left breast mass in a heatmap by a faint blue/green colour hue and displayed abnormality scoring of 56% suspicion of malignancy. Such mass was diagnosed BI-RADS 4c on conventional breast imaging and malignancy was highly suspected. The mass proved to be benign phylloids by tissue core biopsy and the suspicious secondary signs of: focal tissue edema, thickened Cooper's ligament and dermal thickening were likely due to inflammatory changes from the rapid rate of growth. AI abnormality scoring of more than 49.5% (i.e. 56%) suggested phylloids rather than fibro-adenoma and benign phylloids not malignant one since the value was less than 90%

had perforated capsule with finger like projections that is why wide excision surgical procedure is required unlike FA where only simple excision may be needed [10]. Even with the aid of the ultrasound, unless there is a history of rapid increase in the size of the mass; their distinction from each other in many situations does present a dilemma.

In the current work, the performance of the artificial intelligence was studied with regard the diagnosis and discrimination between the PT and FA that presented as masses on mammograms. Ultrasound evaluation was included since it is the descriptive modality which in most situations provides the proper distinction between the variable breast pathologies and sometimes the margins of the masses may be obscured by the glandular tissue on the mammograms.

Previous studies that had focused on the mammograms and AI were concerned mainly with the performance of the AI as a stand-alone screening strategy or as a complementary reading tool to mammogram for the detection

of breast cancer in the screening practice not as a tool of disease discrimination [11].

A recent study by Mansour et al. [12] studied the diagnostic performance of the AI- scanned mammograms in correlation with the traditional used conventional breast imaging modalities (the mammogram and the ultrasound) with regard the different breast entities. Another study assessed the impact of the ultrasound artificial intelligence on the differentiation between benign and malignant breast lesions of BI-RADS 4A [13].

Also, there was specific work that considered the potential role of artificial intelligence in the distinction between phyllodes and fibro-adenoma with regard the AI-aided ultrasound [14] or the whole-slide images in core biopsies [15].

The current work was also considered with such differentiation between phyllodes and fibro-adenoma, however it is a leading work to discuss the assessment using the AI- scanned mammograms.

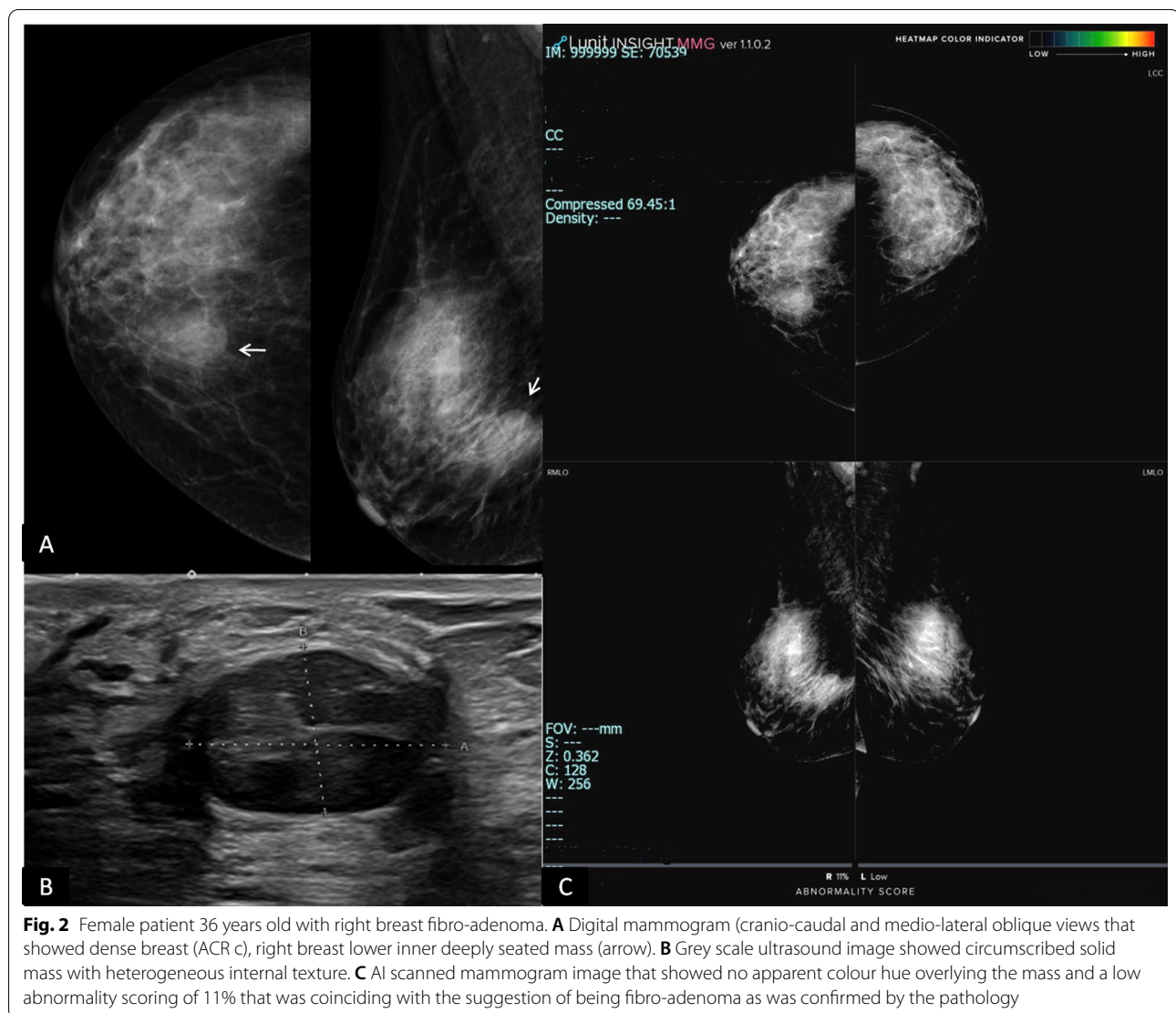


Fig. 2 Female patient 36 years old with right breast fibro-adenoma. **A** Digital mammogram (cranio-caudal and medio-lateral oblique views that showed dense breast (ACR c), right breast lower inner deeply seated mass (arrow). **B** Grey scale ultrasound image showed circumscribed solid mass with heterogeneous internal texture. **C** AI scanned mammogram image that showed no apparent colour hue overlying the mass and a low abnormality scoring of 11% that was coinciding with the suggestion of being fibro-adenoma as was confirmed by the pathology

Re-scanning of the mammograms was done through AI in the form of focal color to target the breast mass (already proved as phylloids or fibro-adenoma) on the mammogram that was supported with an auto-applied abnormality scoring percentage of these masses. Then, a correlation between this numerical estimate (i.e., the abnormality scoring percentage) and the pathology results was performed.

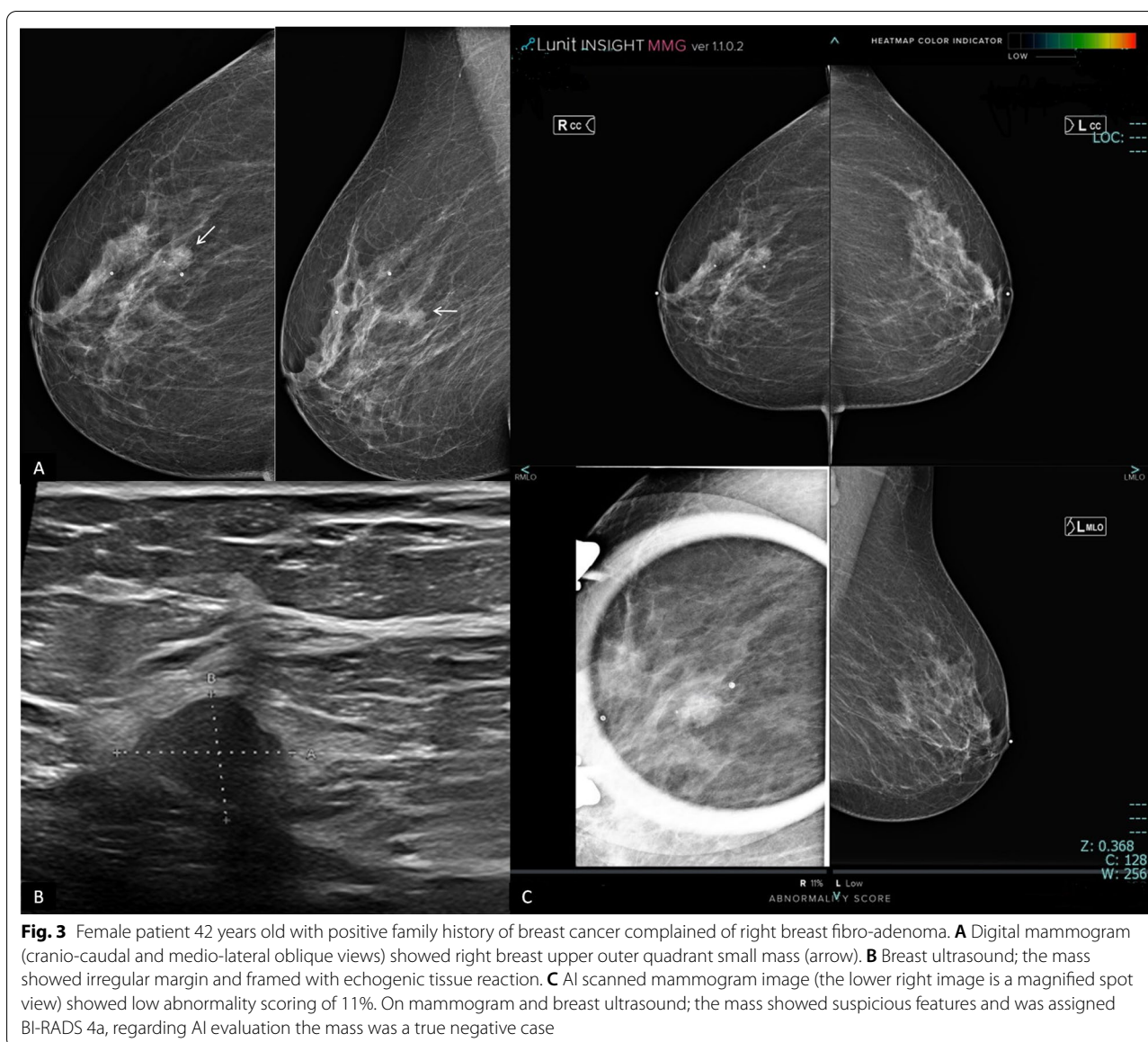
Duman et al. [3], suggested that FA could be differentiated from PT by the shape of the tumor, while other groups found no significant difference in the shape between both tumors which tend to be oval or irregular in shape; phylloides tumors may grow more rapidly than fibroadenomas on follow-up ultrasonography, but they cannot be reliably differentiated by imaging [14, 16, 17].

In this work, PT presented mainly with the rounded / oval shape pattern (61% versus 55% for FA), while the

irregular shape went more with the FA (45% versus 39% for PT).

Lee et al. [18] reviewed in accordance with most findings in the literature; that PT presented predominantly with circumscribed margins. This went in concordance with the present work, as circumscribed margins were more common in PT (75.6% versus 61.9% for the FA). However, they disagreed with Duman et al. [3], who stated that circumscribed margins were significantly more common in FA than in PT.

In the study by Wiratkapun et al. [19], 85% of the included PT was complex/heterogeneous masses, as reported in other studies where FA was commonly presented by homogenous texture [3, 19]. The current study also showed that FA was mainly homogenous (57.2% versus 17.7% for PA).



In this work, when mammograms were scanned with the AI algorithms, correct diagnosis presented in 94.8% ($n=199/210$) FA and 89.6% ($n=147/164$) PT masses Figs. 1 and 2.

Many benign breast diseases show irregular hypo-echoic masses that can mimic carcinoma on ultrasound [20].

Even when masses displayed irregular shape; low abnormality scoring percentage- elicited at the AI scanned mammograms- favor benign nature of the mass and so follow up could be recommend to the patient rather than biopsy or unnecessary surgical removal, Fig. 3.

PT commonly is presented as a rapidly growing mass that could be associated with significant painful erythema and warmth of the overlying breast skin [21]. This is a very misleading feature that can delay the diagnosis and sometimes may suggest malignant pathology. However, the diagnosis of PT rather than FA was easily applicable in the current work when the cut off value of the abnormality scoring was more than 49.5%, Figs. 1 and 4.

Such AI related- probability had a positive impact on upgrading the sensitivity and the specificity of the conventional breast imaging from 75.8% and 59.2% to 89.6% and 94.8% respectively.

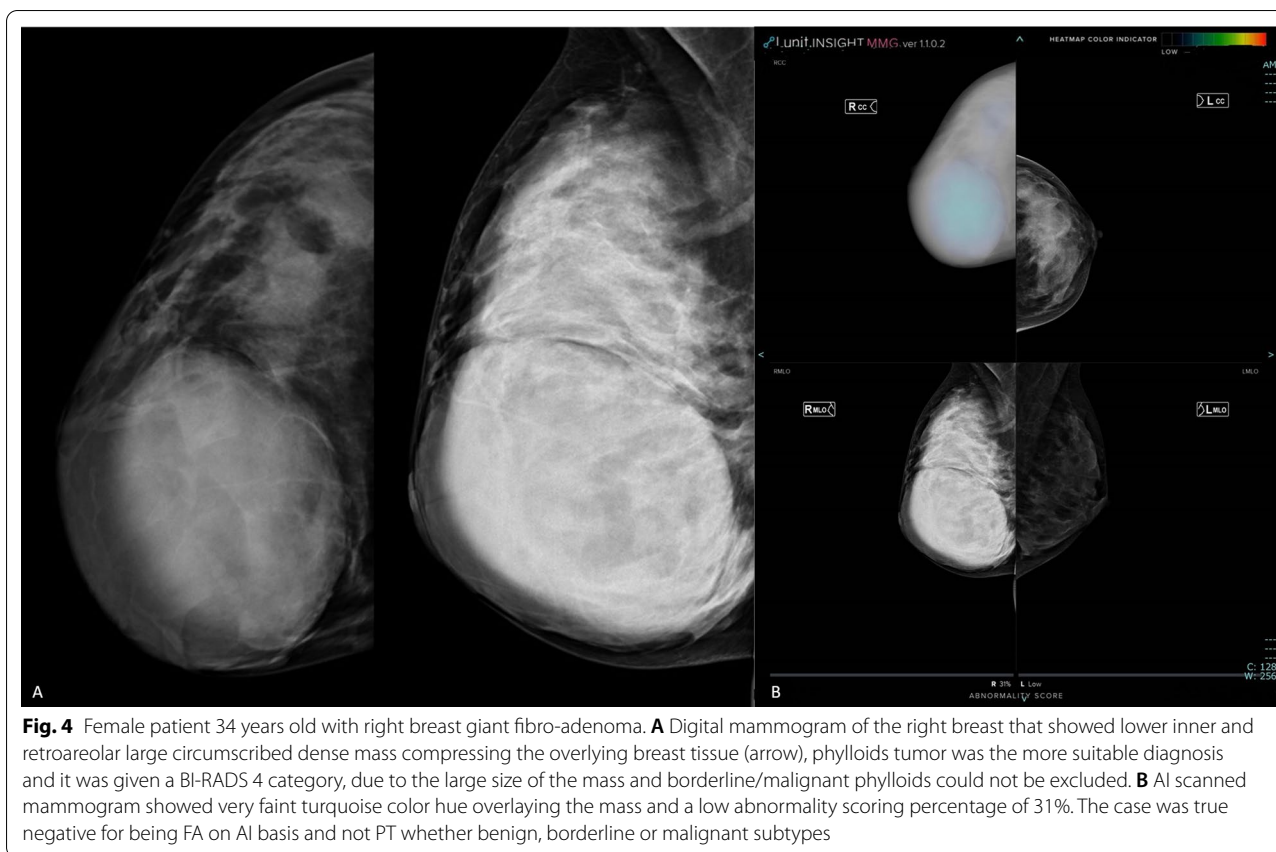


Fig. 4 Female patient 34 years old with right breast giant fibro-adenoma. **A** Digital mammogram of the right breast that showed lower inner and retroareolar large circumscribed dense mass compressing the overlying breast tissue (arrow), phylloids tumor was the more suitable diagnosis and it was given a BI-RADS 4 category, due to the large size of the mass and borderline/malignant phylloids could not be excluded. **B** AI scanned mammogram showed very faint turquoise color hue overlaying the mass and a low abnormality scoring percentage of 31%. The case was true negative for being FA on AI basis and not PT whether benign, borderline or malignant subtypes

Large tumor size at presentation or rapid growth raises the suspicion for a phylloids tumor rather than a fibro-adenoma [22]. Fibro-adenoma could show large size as well, which is usually encountered in pregnant or lactating women as their growth is associated with increase in the estrogen, progesterone, and prolactin hormones and at this situation it is termed giant fibro-adenomas. Giant fibro-adenomas could overlap with borderline or malignant phyllodes [23].

In the current study, the accurate diagnosis of large masses more than 5 cm was helped using the abnormality scoring of the AI algorithm into PT or FA, Figs. 1 and 4.

Benign, borderline, and malignant PT of the breast have similar imaging features; some MRI findings can be used to determine the risk of malignancy which include non-circumscribed margins, peri-tumor edema, and low signal intensity on T2-weighted images. On basis of the conventional breast imaging, this task is a challenge [24].

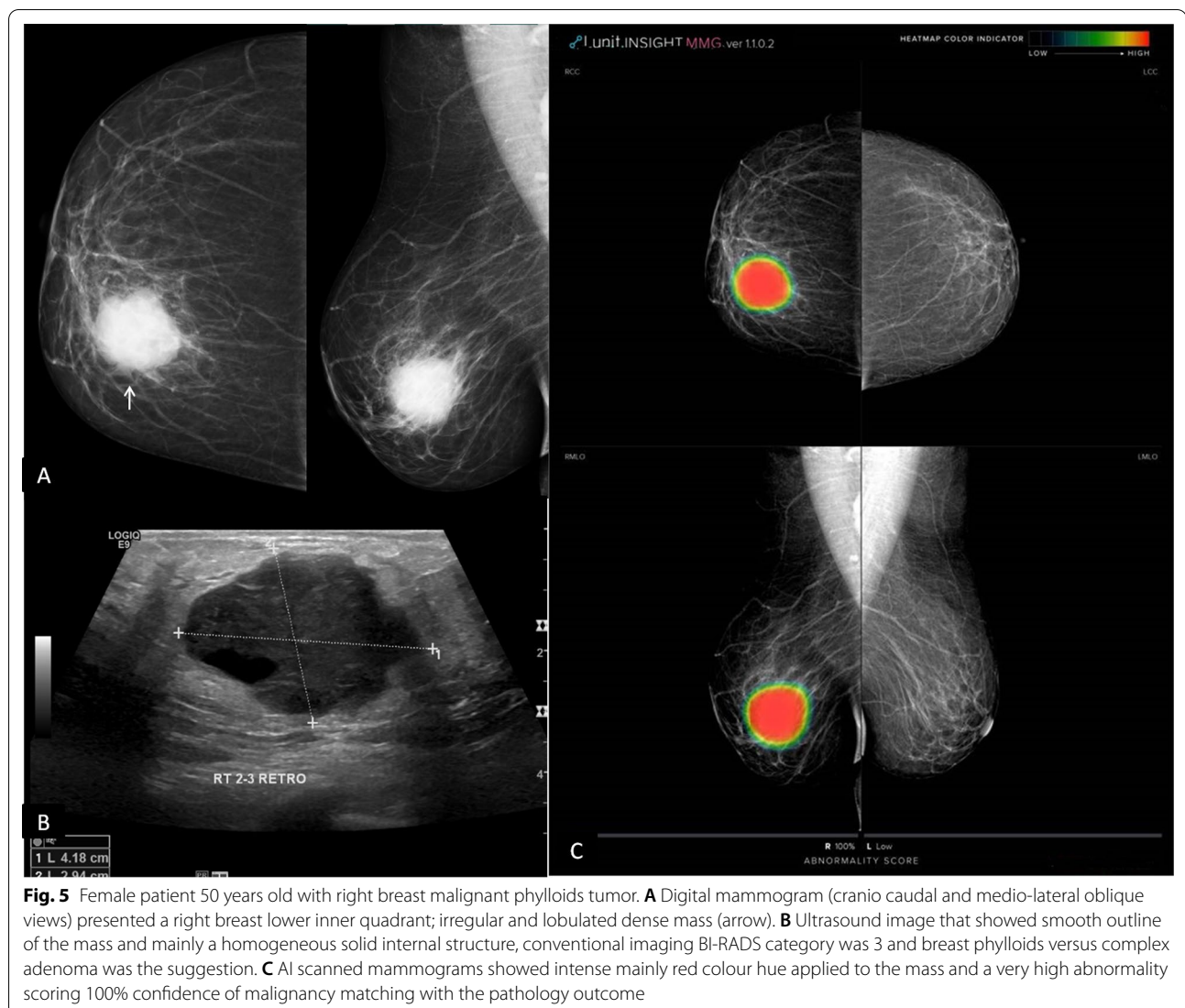
A study performed in 2012, by Dheeba and Selvi [25] showed one of the highest sensitivity (96.9%) and specificity (92.9%) for the proposed AI algorithm in the detection of the cancer in the mammograms. In 2021; the study by Mansour and co-authors [12] showed a sensitivity of

96.8% and a specificity of 90.1% in the discrimination between benign and malignant breast lesions.

Distinction of malignant PA masses that were included in this study was applicable in 100% of these masses ($n = 12/12$) by the aid of the AI- mammogram combination. This was the condition with masses that presented on mammogram with high density, haziness of the tissue surrounding the tumor due to edema, superimposition of a predominantly intense red color hue on the AI images and a correlating high abnormality scoring suspicion of malignancy that was more than 90%, Fig. 5.

According to Stavros et al. [26], the AI- feature analytic algorithms may support subdividing the BI-RADS category 4, thus can prompt the use of BI-RADS-based structured reporting and encourage the reconsider of tissue sampling for these lesions.

This was the condition with some masses who presented by suspicious features on the primary evaluation by the mammogram and ultrasound, yet when these mammograms were scanned by AI, these masses were not overlaid by color hue and eventually the breast was assigned a low scoring percentage of less than 10% (i.e. lesions of low significance) Figs. 1, 3 and 4. In these cases,



if the AI pattern of interpretation were considered in the clinical setting, then biopsy could have been dismissed (especially in case of the small sized masses). Biopsy is warranted if masses presented large size. The interventional procedure in such condition is needed to confirm the absence of malignant potential. The indulgence of the AI findings in the decision of the diagnosis would save the patient from unneeded panic and/or anxiety in large sized benign looking masses.

The heatmap (i.e., color hue) elicited by AI on the scanned mammogram images is used to spot abnormal breast lesions, so it could guide future follow ups or localize lesions that warrant biopsy from those that require further imaging settings. Biopsy is to be considered in case of serial increase in the abnormality scoring percentage of masses under follow up [12].

The current experience showed that masses with low scoring percentage on AI scanned mammograms can then be subjected to interval supervision. Even in case of stationary morphologic features, AI could be considered as a parameter of follow up; in the by monitoring the changes in the intensity of the color hue (i.e., changes from cold light blue or green colors to intense hot colors as yellow, orange, or red) on the mammogram and /or in the value of the abnormality scoring percentage (i.e., increase in the value of the abnormality scoring percentage).

However, precise reports and proper recommendation of the management require continuous supervision of the AI performance by the radiologist [27].

In the current work, AI presented 17 false negative cases that were misdiagnosed as FA instead of PT. That

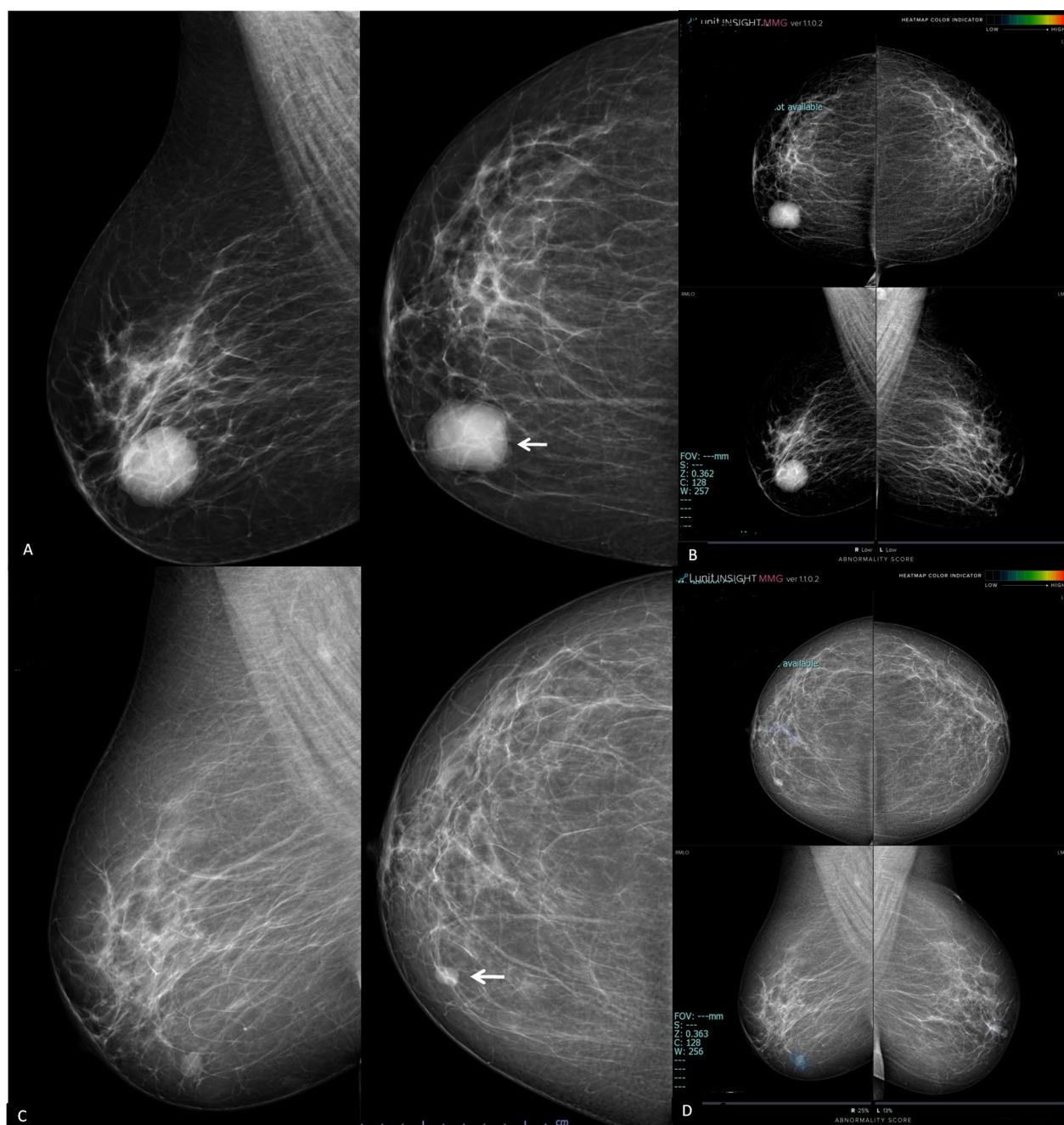


Fig. 6 Female patient 40 years old with right breast benign phylloids tumor. **A** Digital mammogram (cranio caudal and medio-lateral oblique views) that showed a right breast lower inner quadrant rounded shaped dense mass (arrow). **B** AI scanned mammograms showed no colour hue applied to the mass and a low abnormality scoring (i.e., < 10%). The mass was benign looking given BI-RADS 2 category and at this stage, FA was the expected diagnosis. **C** A previous 12 months earlier comparable digital mammogram (that was available on the system) showed a smaller size of the right breast mass. This time the mass displayed faint blue color hue and a low abnormality scoring of 25% on the AI scanned mammogram seen at **(D)**. The mass in **(C)** and **(D)** was classical of being FA and was given BI-RADS 2 category and was subjected to interval follow up instead of biopsy. The increase in the size of the mass was the reason for the application of intervention. The AI missed the clinical history of the patient, which affected the proper evaluation of the AI algorithm and subjected the patient for a while to improper management

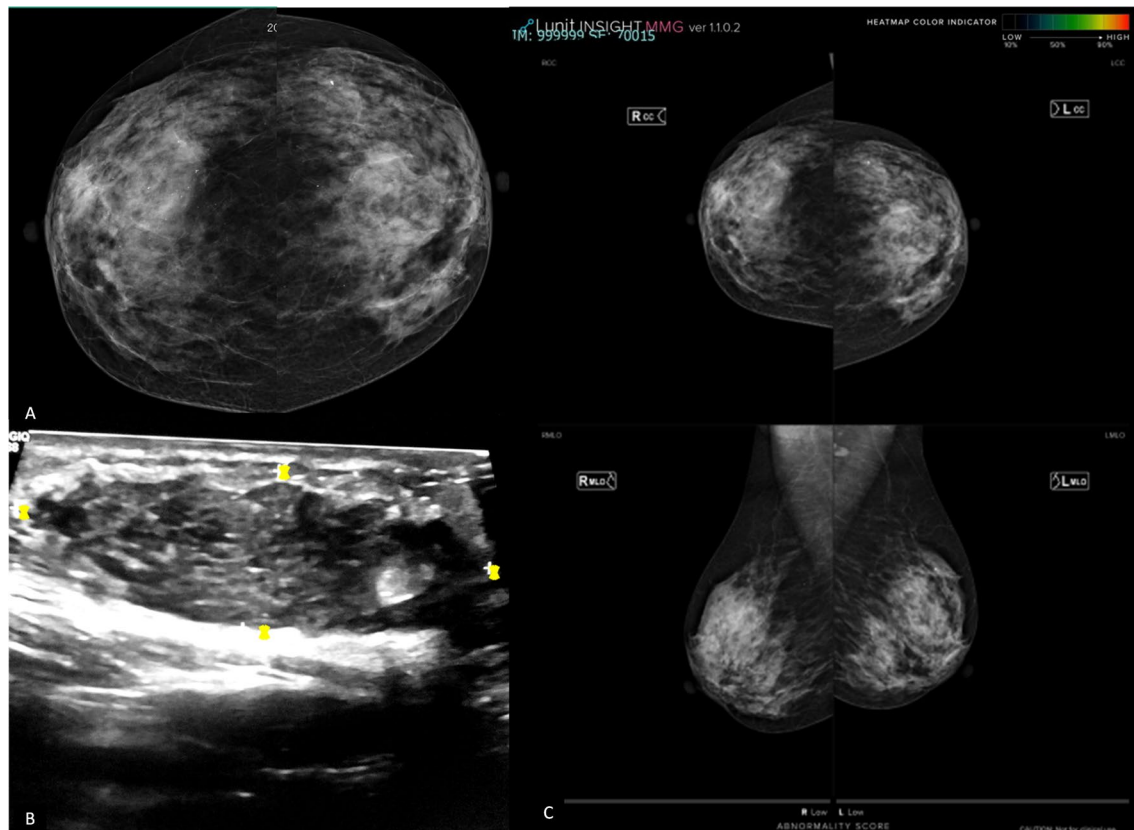


Fig. 7 Female patient 38 years old with right breast peri-canalicular fibro-adenoma. **A** Digital mammogram (cranio-caudal view), that displayed dense breasts (ACR C) and no obvious masses could be detected. **B** Breast ultrasound showed, right breast upper outer quadrant; irregular heterogeneous purely solid mass in plane with the breast tissue. So mammogram was assigned BI-RADS zero and ultrasound was given BI-RADS 4, and biopsy was requested. **C** AI scanned mammogram image showed no localized abnormality by the colour hue and the abnormality scoring for both breasts was less than 10%. The low scoring by the AI supported the benign suggestion of the right breast mass yet, the absence of color hue by the AI and the high breast density (the mass couldn't be visualized by the reader on the mammogram) would have given the breast a wrong BI-RADS score of 1 and consequently an improper management and recommendation

assumption was based on the low abnormality scoring percentage that sometimes presented a value of less than 10%, Fig. 6. The example case: Fig. 6, presented with right breast mass that rapidly increased in size and this was the point that supported the recommendation of the biopsy; otherwise, the mass showed the classic morphologic features of benignity on the baseline as well as the follow up mammograms. Also, the AI scanned images showed a low abnormality scoring that didn't exceed 25%. That's why, it is important to keep direct communication with the patient to be accounted with the clinical history and the circumstances of the breast disease.

In breast densities with ACR a and b; masses are obvious with respect to the breast glandular tissue and so mammograms scanned with AI that showed no color hue demarcation and was given a low abnormality scoring of suspicion (i.e., less than 10%) could be considered as a satisfactory modality of scanning. Yet, in case of breast densities assigned ACR c or d patterns;

further scanning with ultrasound is required so that not to miss masses overlapped by the dense glandular tissue of the breast and be sure about the BI-RADS category whether normal (category 1) or benign (category 2) Fig. 7.

The current experience showed significant correlation between the BI-RADS category that was assigned by the mammogram and ultrasound combination and the abnormality scoring elicited by the AI scanning with regards the categories; "likely benign, BI-RADS 2" (P value < 0.001), "probably malignant, BI-RADS 4" (P value 0.020) and "likely malignant, BI-RADS 5" (P value < 0.001).

There were limitations to the current work: (1) its retrospective nature limited the ability to determine the Doppler ultrasound details of all the included masses. Resistive index value has not been reported previously and moreover it was not possible to determine whether the distribution vascularity was central or peripheral.

(2) Clinical data and demographic findings of the cases were not analyzed, as this study was based only on imaging findings.

Conclusion

Artificial intelligence-aided mammograms could be used as method of distinction between PT from FA detected on sono-mammogram. The color hue and the quantification of the abnormality scoring percentage could be used as a one setting method for specification and so guide clinicians in their decision of conservative management or the choice of the surgical procedure.

Abbreviations

AI: Artificial intelligence; PT: Phylloids tumor; FA: Fibroadenoma; BI-RADS: Breast Imaging Reporting and Data System; CNN: Convolutional neural networks; SD: Standard deviation; PPV: Positive predictive value; NPV: Negative predictive value; ACR: American college of Radiology.

Authors' contributions

SA is the guarantor of integrity of the entire study. MA and MS contributed to the study concepts and design. MS, MA, and HC contributed to the literature research. MS, MA, HC and YA contributed to the clinical studies. MS and YA contributed to the experimental studies/data analysis. MS and MA contributed to the statistical analysis. MS, HC, and YA contributed to the manuscript preparation. MS, HC, and YA contributed to the manuscript editing. All authors have read and approved the final manuscript.

Funding

No source of funding.

Availability of data and materials

The corresponding author is responsible for sending the used data and materials upon request.

Declarations

Ethics approval and consent to participate

The study was approved by the ethical committee of the Radiology Department of El Kasr ElAiny Hospital, Cairo University which is an academic governmental supported highly specialized multidisciplinary Hospital. A waiver of informed consent was obtained.

Consent for publication

All patients included in this research were legible, above 16 years of age. The study was waived to get written informed consent from the included patients.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Women's Imaging Unit, Radiology Department, Kasr ElAiny Hospital, Cairo University, ElManial, Cairo, Egypt. ²AI Informatics, FUJIFILM Medical Imaging and IT Solutions MEA, Dubai, United Arab Emirates. ³Kasr ElAiny Hospital, Cairo University, Cairo, Egypt.

Received: 14 October 2021 Accepted: 21 February 2022

Published online: 01 March 2022

References

- Hasdemir S, Tolunay S, Özşen M, Gökgöz MS (2020) Phylloides tumor of the breast: a clinicopathological evaluation of 55 cases. *Eur J Breast Health* 16(1):32–38
- Stoffel E, Becker AS, Wurnig MC, Marcon M, Ghafoor S, Berger N et al (2018) Distinction between phylloides tumor and fibroadenoma in breast ultrasound using deep learning image analysis. *Eur J Radiol Open* 5:165–170
- Duman L, Gezer NS, Balci P, Altay C, Başara I, Durak MG et al (2016) Differentiation between phylloids tumors and fibroadenomas based on mammographic sonographic and MRI features. *Breast Care* 11:123–127
- Geras KJ, Mann RM, Moy L (2019) Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology* 293:246–259
- Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Panykh OS et al (2018) Current applications and future impact of machine learning in radiology. *Radiology* 288(2):318–328
- Sechopoulos I, Teuwen J, Manna R (2021) Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: state of the art. *Semin Cancer Biol* 72:214–225
- Chiwome L, Okojie OM, Rahman AKM, Javed F, Pousette Hamid P (2020) Artificial intelligence: is it armageddon for breast radiologists? *Cureus* 12(6):e8923
- D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA et al (2013) ACR BI-RADS® Atlas, breast imaging reporting and data system. American College of Radiology, Reston, VA
- Plaza MJ, Swintelski C, Yaziji H, Torres-Salichs M, Esserman LE, LE, (2015) Phylloides tumor: review of key imaging characteristics. *Breast Dis* 35(2):79–86
- Limaem F, Kashyap S (2022) Phylloides tumor of the breast. [Updated 2021 Nov 5]. In: StatPearls. StatPearls Publishing, Treasure Island, FL. <https://www.ncbi.nlm.nih.gov/books/NBK541138/>. Accessed 4 Feb 2022
- Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A et al (2021) Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ* 374:n1872
- Mansour S, Kamal R, Hashem L, AlKalaawy B (2021) Can artificial intelligence replace ultrasound as a complementary tool to mammogram for the diagnosis of the breast cancer? *Br J Radiol* 94:20210820
- Niu S, Huang J, Li J, Liu X, Wang D, Zhang R et al (2020) Application of ultrasound artificial intelligence in the differential diagnosis between benign and malignant breast lesions of BI-RADS 4A. *BMC Cancer* 20(20):959
- Niu S, Huang J, Li J, Liu X, Wang D, Wang Y et al (2021) Differential diagnosis between small breast phylloides tumors and fibroadenomas using artificial intelligence and ultrasound data. *Quant Imaging Med Surg* 11(5):2052–2061
- Cheng CL, Md Nasir ND, Ng GJZ, Chua KWJ, Li Y, Rodrigues J (2021) Artificial intelligence modelling in differentiating core biopsies of fibroadenoma from phylloides tumor. *Lab Invest*. Epub ahead of print
- Ng CCY, Md Nasir ND, Loke BN, Tay TKY, Thihe AA, Rajasegaran V et al (2021) Genetic differences between benign phylloides tumors and fibroadenomas revealed through targeted next generation sequencing. *Mod Pathol* 34:1320–1332
- Kalambo M, Adrada BE, Adeyefa MM, Krishnamurthy S, Hess K, Carkaci S et al (2018) Phylloides tumor of the breast: ultrasound-pathology correlation. *Am J Roentgenol* 210:W173–179
- Lee JS, Chang J, Hagemann IS, Bennett DL (2021) Malignant phylloides tumor: imaging features with histopathologic correlation. *J Breast Imaging* 3(6):703–711
- Wiratkapun C, Piyapan P, Lertsithichai P, Larbcharoensub N (2014) Fibroadenoma versus phylloides tumor: distinguishing factors in patients diagnosed with fibroepithelial lesions after a core needle biopsy. *Diagn Interv Radiol* 2014(20):27–33
- Kim YR, Kim HS, Kim H (2015) Are Irregular Hypoechoic Breast Masses on Ultrasound Always Malignancies? A Pictorial Essay *Korean J Radiol* 16(6):1266–1275
- Testa I, Salvatori C, Prestipino M, Laurenti ME, Gerli P, Cara GD et al (2018) Inflamed phylloides tumour in a girl: a challenging diagnosis in paediatric breast lesions. *Int J Environ Res Public Health* 15(5):959
- Demian GA, Fayaz S, Eissa HE, Nazmy N, Samir S, George T (2016) Phylloides tumors of the breast: analysis of 35 cases from a single institution. *J Egypt Natl Cancer Inst* 28(4):243–248

23. Meng X, Yamanouchi K, Kuba S, Sakimura C, Morita M, Matsuguma K et al (2019) Giant fibroadenoma of the breast: a rare case in a mature woman. *Int J Surg Case Rep* 63:36–39
24. Fede ABS, Pereira Souza R, Doi M, De Brot M, Aparecida Bueno de Toledo Osorio C, Rocha Melo Gondim G et al (2021) Malignant phyllodes tumor of the breast: a practice review. *Clin Pract* 11(2):205–215
25. Dheeba J, Selvi TS (2012) An improved decision support system for detection of lesions in mammograms using differential evolution optimized wavelet neural network. *J Med Syst* 36(5):3223–3232
26. Stavros AT, Freitas AG, deMello GGN, Barke L, McDonald D, Kaske T et al (2017) Ultrasound positive predictive values by BI-RADS categories 3–5 for solid masses: an independent reader study. *Eur Radiol* 27:4307–4315
27. Mendelson E (2019) Artificial intelligence in breast imaging: potentials and limitations. *AJR Am J Roentgenol* 212(2):293–299

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
