BJR

https://doi.org/10.1259/bjr.20210820

Received: 07 July 2021 | Revised: 17 August 2021 | Accepted: 21 September 2021

Cite this article as:
Mansour S, Kamal R, Hashem L, AlKalaawy B. Can artificial intelligence replace ultrasound as a complementary tool to mammogram for the diagnosis of the breast cancer?. *Br J Radiol* 2021; **94**: 20210820.

FULL PAPER

Can artificial intelligence replace ultrasound as a complementary tool to mammogram for the diagnosis of the breast cancer?

[1,2]**SAHAR MANSOUR, MD, [1,2]RASHA KAMAL, [1,2]LAMIAA HASHEM and [1,2]BASMA ALKALAAWY**

[1]Women's Imaging Unit – Kasr El Ainy Hospital- Cairo University, Giza, Egypt
[2]Department of Radiology, Baheya Foundation for Early Detection and Treatment of Breast Cancer, Giza, Egypt

Address correspondence to: Dr Sahar Mansour
E-mail: *sahar_mnsr@yahoo.com*

Objective: To study the impact of artificial intelligence (AI) on the performance of mammogram with regard to the classification of the detected breast lesions in correlation to ultrasound-aided mammograms.
Methods: Ethics committee approval was obtained in this prospective analysis. The study included 2000 mammograms. The mammograms were interpreted by the radiologists and breast ultrasound was performed for all cases. The Breast Imaging Reporting and Data System (BI-RADS) score was applied regarding the combined evaluation of the mammogram and the ultrasound modalities. Each breast side was individually assessed with the aid of AI scanning in the form of targeted heat-map and then, a probability of malignancy (abnormality scoring percentage) was obtained. Operative and the histopathology data were the standard of reference.
Results: Normal assigned cases (BI-RADS 1) with no lesions were excluded from the statistical evaluation. The study included 538 benign and 642 malignant breast lesions (n = 1180, 59%). BI-RADS categories for the breast lesions with regard to the combined evaluation of the digital mammogram and ultrasound were assigned BI-RADS 2 (Benign) in 385 lesions with AI median value of the abnormality scoring percentage of 10 (n = 385/1180, 32.6%), and BI-RADS 5 (malignant) in 471, that had showed median percentage AI value of 88 (n = 471/1180, 39.9%). AI abnormality scoring of 59% yielded a sensitivity of 96.8% and specificity of 90.1% in the discrimination of the breast lesions detected on the included mammograms.
Conclusion: AI could be considered as an optional primary reliable complementary tool to the digital mammogram for the evaluation of the breast lesions. The color hue and the abnormality scoring percentage presented a credible method for the detection and discrimination of breast cancer of near accuracy to the breast ultrasound. So consequently, AI- mammogram combination could be used as a one setting method to discriminate between cases that require further imaging or biopsy from those that need only time interval follows up.
Advances in knowledge: Recently, the indulgence of AI in the work-up of breast cancer was concerned. AI noted as a screening strategy for the detection of breast cancer. In the current work, the performance of AI was studied with regard to the diagnosis not just the detection of breast cancer in the mammographic-detected breast lesions. The evaluation was concerned with AI as a possible complementary reading tool to mammogram and included the qualitative assessment of the color hue and the quantitative integration of the abnormality scoring percentage.

INTRODUCTION

Mammography sometimes is not the suitable choice for breast imaging especially when the glandular tissue showed the dense pattern.[1]

Tissue overlap that may display between the normal tissue and tumors of the breast limits the sensitivity of the mammogram. Moreover, if the tumors were easily seen, the actual extend of the disease may not be clearly demonstrated.[2]

Recently, the possible use of artificial intelligence (AI) is increasingly studied in the screening of breast cancer. The advances in deep learning will likely present a valuable participation to upgrade the performance of digital mammogram for breast imaging.[3]

AI is expected to be potentially applied for the diagnosis of the breast cancer, determination of its extent within the breast tissue and the interpretation of the pathology.[4]

Ongoing studies suggested the AI algorithm is about to reach a high level of performance in the context of the detection of the breast cancer at the screening mammography.[5]

The purpose of this work is to perform a prospective study on mammogram scanned with AI algorithm and evaluate its impact on the detection and/or discrimination of cancer in correlation with mammogram aided with complementary ultrasound.

PATIENTS AND METHODS

Patients

The current work is an ethics committee approved prospective analysis that evaluated 1000 patients with total of 2000 mammograms from August 2019 till September 2020. All patients enrolled in the study have given an informed consent.

Inclusion criteria: cases eligible for screening mammogram (45 years or older) and patients more than 30 years of age with clinical complaints.

Exclusion criteria: patients with known breast cancer under chemotherapy follow-up or performed conservative breast surgery. Patients with biopsy proven breast cancer and patients who are not eligible to do screening mammogram as: Cases less than 30 years of age and pregnant patients

Methods

Each patient was imaged with digital mammography, (manufacture: Amulet Innovality, Fujifilm Gobal company, Japan). Mammography machines are supported with a “Bellus” workstation of resolution five megapixels. The breasts were imaged in the craniocaudal and mediolateral oblique views.

Included breasts were then re-scanned by AI software (Lunit INSIGHT MMG, v. 2019) for Fujifilm digital mammography system which was designed to support deep learning reading of the mammogram images through a provider developed algorithms.

The AI algorithm was developed on the basis of deep convolutional neural networks (CNNs). The algorithm training consists of two stages: patch-level training from scratch for learning low-level features (Stage 1), followed by image-level fine tuning from the Stage 1 model for learning high-level context (Stage 2). Only lesion-annotated mammograms were used in Stage 1 (fully supervised), whereas all mammograms were used in Stage 2 (semi-supervised). Batch-instance normalization[6] and a deconvolution module[7] were additionally adopted to overcome variance of pixel-level characteristics and increase of false positives, respectively. For an input mammogram image (*i.e.* one of the four views), the AI algorithm provides pixel-level abnormality scores as a heatmap and a representative abnormality score, which is the maximum of the pixel-level abnormality scores. The abnormality scores are floating-point values between 0 and 1.

Based on a per-image analysis of the algorithm, the resulting diagnostic support software 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 to detect breast lesions.

Included cases were also subjected to complementary ultrasound examination (HS60 Samsung ultrasound, Korea, 2019) to support the BI-RADS category provided by the digital mammogram and the detected lesions seen at the snapshot images presented via the AI system.

Image analysis

Three readers (25 years experience in breast imaging each) assessed all mammograms in individual sessions.

Mammograms were given the five BI-RADS categories,[8] where BI-RADS category number 1 was given for the normal appearing mammograms. Abnormalities detected on the mammograms included: mass, focal asymmetry, focal distortion or calcifications.

The breast lesions detected on the mammogram was then re-evaluated by ultrasound regarding; texture patterns (solid or cystic or complex), long axis orientation (in plane with the breast tissue or vertical infiltration) and the presence of secondary suspicious signs of breast disease as interstitial edematous changes and dermal thickening.

The BI-RADS category was applied after the inclusion of the findings of the complementary breast ultrasound to those of the mammography.

The AI software provided a localization of detected abnormalities that was categorized by the system algorithm. AI interpretation was performed regarding qualitataive and quantitative criteria. (I) Qualitative criteria: data visualization that represents the magnitude of abnormality in the form of a color hue “heat map”. The color hue is a mix of different colored patches not a pure pigment (*i.e.* not distinctive colors). These colors mix marks the breast lesion on the mammogram images. The colors indulged ranged from the light turquoise blue to the intense red. Such range of colors was superimposed on each other; where the degree of hotness of the color was increased in correlation with the increase in the value of the abnormality scoring percentage which is a reflection to the confidence of suspicion of malignancy.

(II) Quantitative criteria: numerical estimate of the degree of confidence for the suspicion of malignancy “abnormality scoring” that ranges from 1 to 100, with the value 100 representing the highest level of suspicion.

The study included only the pathologically proven cases. The pathology data obtained via ultrasound-guided core biopsy/surgery was the standard of reference.

Statistical analysis

The collected data were coded, tabulated, and statistically analyzed using a statistical package (SPSS v. 20). Descriptive statistics were performed for quantitative parametric data as mean ± SD, whereas they were performed for qualitative data as number and percentage.

Standard diagnostic indices including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and diagnostic accuracy were calculated. Comparisons of sensitivity or specificity between the combined mammography and ultrasound and the AI performance were performed. Comparison between categorical data was performed using χ^2 test. The level of statistical significance (p-value) for all tests was set at 0.05. The p-value is a statistical measure for the probability that the results observed in a study could have occurred by chance.

Results

500 patients were included in the current work (590 breasts/ mammograms); their age ranged from 30 to 63 years old. Mean age was 49 ± 10.6 years (mean ± SD).

Normal assigned cases (BI-RADS 1) were excluded from the statistical evaluation. Cases for analysis were those with detected abnormalities (*i.e.* BI-RADS 2, 3, 4 and 5) on conventional imaging (n = 1180): The study included 538 benign and 642 malignant breast lesions

Positive cases were those with BI-RADS 4 and 5 categorizes (n = 678/1180, 57%).

Table 1 displayed the various types of breast lesions included in the study. The benign lesions included fibroadenoma in 43.5% (n = 234), fibrocystic disease in

16.7% (n = 90), infectious mastitis in 11.2% (n = 60), fat necrosis in 7.4% (n = 40), atypical intraductal epithelial hyperplasia in 6.7% (n = 36), granulomatous mastitis in 5.2% (n = 28), fibroadenolipoma in 3.7% (n = 20), nodular sclerosis in 3% (n = 16), stromal fibrosis/fibrous mastopathy in 2.6% (n = 14).

The malignant lesions included invasive ductal carcinoma in 82.4% (n = 529), ductal carcinoma *in situ* in 10.3% (n = 66), invasive lobular carcinoma in 3.2% (n = 21), invasive papillary carcinoma in 1.1% (n = 7), malignant phylloids tumor in 0.5% (n = 3), mucinous carcinoma in 1.4% (n = 9), medullary carcinoma in 0.6% (n = 4) and inflammatory carcinomatosis 0.4% (n = 3).

Table 1. Variety of breast lesions that was included in the study

Pathology		Number (Percentage)
Benign	Fibroadenoma	234 (43.5%)
	Fibrocystic disease	90 (16.7%)
	Infectious mastitis	60 (11.2 %)
	Fat necrosis	40 (7.4%)
	Atypical intraductal epithelial hyperplasia	36 (6.7%)
	Granulomatous mastitis	28 (5.2%)
	Fibroadenolipoma	20 (3.7%)
	Nodular sclerosis	16 (3%)
	Stromal fibrosis/fibrous mastopathy	14 (2.6%)
Malignant	Invasive ductal carcinoma	532 (83%)
	Ductal carcinoma *in situ*	66 (10.3%)
	Invasive lobular carcinoma	21 (3.2%)
	Invasive papillary carcinoma	7 (1.1%)
	Mucinous carcinoma	9 (1.4%)
	Medullary carcinoma	4 (0.6%)
	Inflammtory carcinomatosis	3 (0.4%)

The histological grade of these breast invasive and non-invasive carcinomas were Grade I in 16% (n = 85), Grade II in 69% (n = 367) and Grade III in 15% (n = 80).

On mammogram, breast lesions presented by masses in 65.4% (n = 771), asymmetrical density and distortion in 18% (n = 213), asymmetry without distortion in 10.6% (n = 125) and calcifications in 6% (n = 71).

On ultrasound lesions abnormalities represented focal lesions of solid texture in 57% (n = 672), cystic pattern in 20% (n = 236),

Table 2. The correlation between the AI assigned percentage of the abnormality scoring and the BI-RADS category of the included breast lesions

AI abnormality score *BI-RADS category	No.	Median	First quartile	Third quartile	Inter quartile range	p value
Likely benign 2	385	10	10	15	5	<0.001
Probably benign 3	117	14	10	30	20	0.741
Probably malignant 4	207	39	17	74	57	0.020
Likely malignant 5	471	88	77	93	16	<0.001

AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System.

Table 3. The correlation between the pathology outcome (benign *vs* malignant) of the included breast lesions and the AI assigned percentage of the abnormality scoring

Pathology *AI abnormality score	No.	Median	First quartile	Third quartile	Inter quartile range
Benign	538	25	18	37	19
Malignant	642	88	78	94	16

AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System.

complex pattern in 17% (*n* = 200). The long axis of the detected lesions were in-plane with the breast tissue in 23% (*n* = 274) and presented with vertical infiltrative form in 51% (*n* = 598). No focal lesions were detected on ultrasound in 6% (*n* = 72), where the presented findings were only edematous changes and dermal thickening.

The BI-RADS categories for the breast lesions as regards the combined evaluation of the digital mammogram and ultrasound were assigned BI-RADS 2 (Benign) in 385 lesions with AI median value of the abnormality scoring percentage of 10, (*n* = 385/1180, 32.6%), BI-RADS 3 (Probably benign) in 117 lesions and median AI % value of 14 (*n* = 117/1180, 10%), BI-RADS 4 (probably malignant) in 207 lesions, median AI % value equals to 39 (*n* = 207/1180, 17.5%) and BI-RADS 5 (malignant) in 471, that had showed median percentage AI value of 88 (*n* = 471/1180, 39.9%)

Table 2 displayed the correlation between the BI-RADS category of the included breast lesions (combined evaluation of the mammogram and the ultrasound assessment) and the AI assigned percentage of the scoring abnormality. The outcome of benign pathology correlated with the AI percentage and presented by interquartile range and median values of 19 and 25 respectively, while for the malignant breast lesions the values were 16 and 88.

Table 3 defined the correlation between the AI scoring of abnormality and the pathology outcome (benign *vs* malignant). To determine a cut-off value for the abnormality scoring percentage that can suggest benign *vs* malignant nature breast lesion, the point on ROC curve was used. AI abnormality scoring of 59% yielded a sensitivity of 96.8% and specificity of 90.1% in the discrimination of the breast lesions detected on the included mammograms.

In the current work, conventional breast imaging - presented by digital mammography and breast ultrasound - displayed 9 false negative and 45 false positive breast lesions, while AI displayed 20 false negative and 53 false positive when correlated with the pathology outcome.

Table 4 shows sensitivity, specificity, accuracy, PPV, and NPV of conventional breast imaging (digital mammogram and ultrasound) and mammograms scanned with AI abnormality scoring.

DISCUSSION

Early detection of breast cancer carries high potentials of successful treatment, that why it is very important to choose the appropriate method for screening of the earliest signs of the cancer.[9]

The screening of breast cancer by mammogram is more efficient with the addition of ultrasound examination especially in females of elevated risk of developing breast cancer and even in an early stage.[10]

Recently, the indulgence of AI in the work-up for the detection of cancer and the probability that it could help the diagnosis, the choice of the treatment, the patient's outcome, and the workflow areas in the field of radiology in general and breast imaging especially has been reviewed earlier in the literature.[11–16]

Yet, the performance of the AI should be monitored by the radiologist in order to prevent the element of an improper management recommendation, incorrect report and the most important to eliminate the possibility of the miscommunication to the patient.[17]

Previous work that focused on the mammograms and AI was concerned with the performance of AI as a standalone screening strategy or as a complementary reading tool to mammogram for the detection of breast cancer.

In the current work, the performance of the AI was studied with regard the to the diagnosis not just the detection of breast cancer in the mammographic-detected breast lesions. The pattern and

Table 4. The statistical indices of the conventional breast imaging (digital mammogram and ultrasound) and those of the AI scanned mammograms with regards the discrimination of the breast lesions

Modality	Sensitivity	Specificity	PPV	NPV	Accuracy	LHR +ve	LHR -ve
Conventional imaging (mammogram and ultrasound)	98.6%	91.6%	93.3%	98.2%	95.4%	1.0883	1.0655
AI scanned mammograms	96.8%	90.1%	92.1%	96%	93.8%	1.0864	1.0633

AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System; LHR, Likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

Figure 1. Female patient 30 years old with left breast upper outer quadrant invasive ductal carcinoma. A, bilateral digital mammogram (craniocaudal and mediolateral oblique views that showed dense breast (ACR c), left breast upper outer focal area of increased density (arrow). B, breast ultrasound (color-coded) that displayed irregular solid mass at the site of compliant with increased vascularity. C, AI scanned mammogram image that marked the left breast upper outer quadrant disease in a heatmap; the disease seen marked by an intense color hue larger than the size displayed by the mammogram/ultrasound and displayed abnormality scoring of 99% confidence of malignancy. D, digital breast tomosynthesis (as another tool of breast imaging) exhibited left breast mass (arrow) and moreover an additional nearby partly circumscribed lesion (arrow), that was not obvious at the mammogram or the ultrasound scanning. Such lesion turned out to be a malignant intramammary sentinel lymph node at the pathology report after surgery. AI not just located the cancer, yet moreover showed the full extent of the disease as confirmed by the operative data. AI, artificial intelligence.

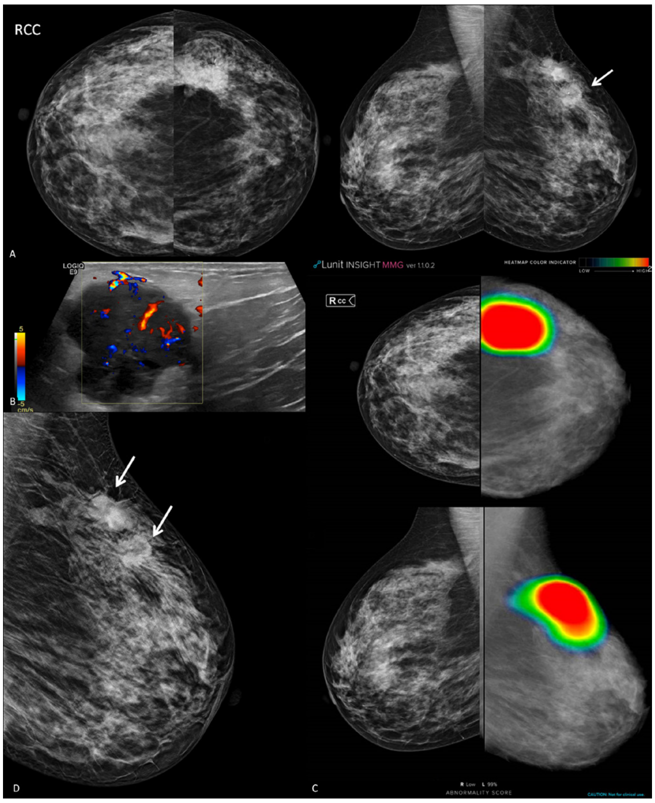
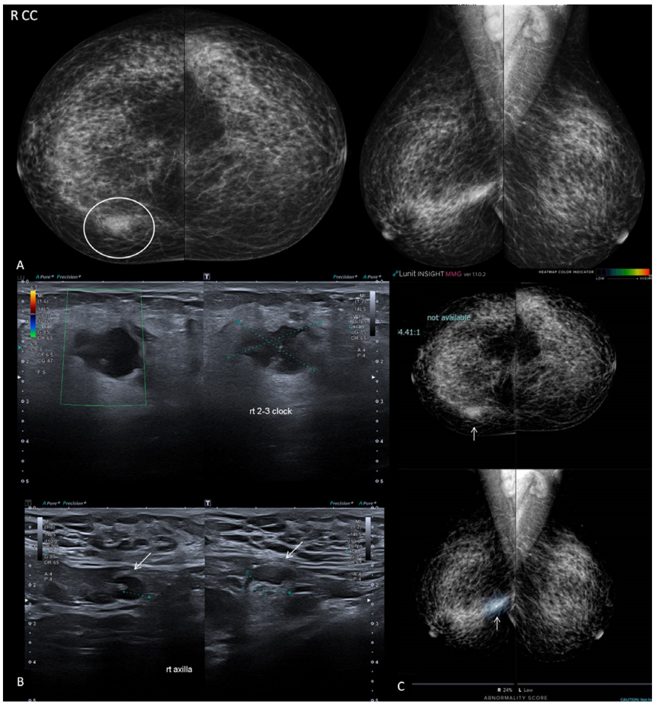


Figure 2. Female patient 36 years old with right breast granulomatous mastitis. A, bilateral digital mammogram (craniocaudal and mediolateral oblique views that showed dense breast (ACR c), right breast lower inner deeply seated focal ill-defined area of asymmetrical density (circle). B, breast ultrasound that showed suspicious ill-defined heterogeneous mainly hypoechoic lesion (upper row; grayscale left image and color coded right one). Lower row showed average sized lymph nodes yet looked indeterminate with asymmetrical cortical thickening of 0.3–0.4 cm and eccentric fatty hilum C, AI scanned mammogram image that marked the left breast abnormality with a faint light blue color hue (arrow) and a low abnormality scoring of 24% that was not coinciding with the suspicious morphology that was presented by the conventional breast imaging. The case was assigned BI-RADS 4 by the conventional imaging and so was a false positive. With respect to the AI; the presented heatmap and the low scoring of suspicion matched with the benign outcome of the pathology. AI, artificial intelligence.



intensity of the focal color in the form of heat map was considered for the qualitative evaluation, and thus localization of the abnormality in addition to the integration of the value of the abnormality scoring in order to find out if there is a correlation between this numerical estimate (*i.e.* confidence of malignancy suspicion) and the nature of the detected lesions.

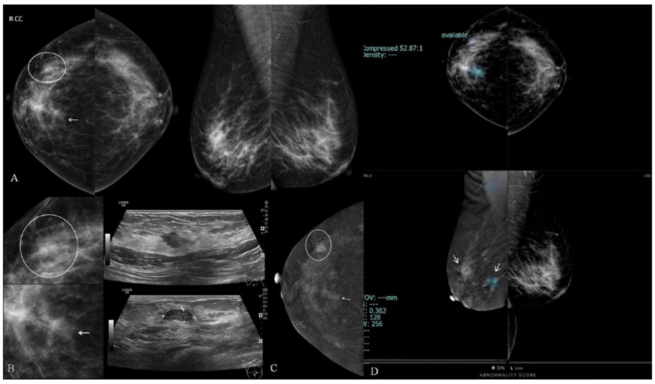
Lesions that had low probability of cancer - *i.e.*<10% scoring- showed no marking with the color hue, then gradually with the increase of the scoring a turquoise color could be visualized starting 15% while those presented with intense red will likely be presented with high scoring of >80% score.

The significance of such color hue on the mammogram images is to spot the lesions of concern presented in the included breasts, so that to localize which portion of the breast was correlating with the abnormality scoring that is seen at the bottom of the image and also to guide the follow up or the biopsy in case further imaging settings were required.

The percentage of suspicion of malignancy was then correlated with the pathology outcome.

The study included 538 benign and 642 malignant lesions detected in 1180 mammograms.

Figure 3. Female patient 42 years old with positive family history of breast cancer complained of right breast mastalgia and performed a diagnostic mammogram. A, bilateral digital mammogram (craniocaudal and mediolateral oblique views) of dense breasts (ACR c). Right breast upper outer ill-defined area of increased density (lesion 1-circle) and related clustered microcalcifications. There was ipsilateral deep central partly obscured lesion (lesion 2-arrow). B, breast ultrasound correlates with the aforementioned right breast two lesions. The upper row showed ill-defined hypoechoic lesion highly suspicious of malignancy in correlation to the lesion no. 1. The lower row showed ultrasound image of lesion no.2 which turned out to be a benign looking circumscribed solid mass. C, contrast-enhanced spectral mammogram of the right breast that displayed i) intense heterogeneous enhancement of lesion no.1 and ii) fair homogenous contrast uptake of lesion no. 2 that was comparable to the background contrast uptake of the breast. So, on conventional breast imaging lesion no. 1 was assigned BI-RADS 4 and lesion no. 2 was considered benign. D, AI scanned mammogram image with displayed faintly demarcation of the right breast lesions (arrows) by the heatmap and a low abnormality scoring of the right breast of 33% was displayed. lesion no.1 at the right breast displayed probably malignant morphology and the contrast uptake increased the suspicion of right breast carcinoma yet pathology reported benign lesion that was nodular sclerosis matching with the AI findings and presenting another false-positive case regarding conventional imaging and a specific diagnosis of true negative by the AI.

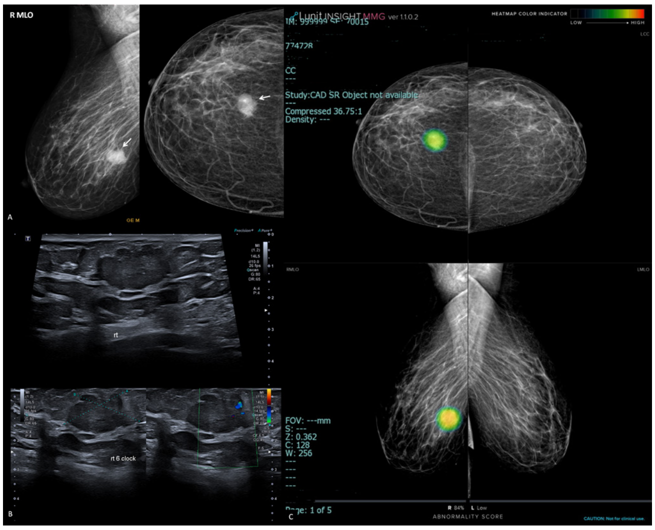


Since mammogram is the modality of screening (*i.e.* detection) and ultrasound is the modality of discrimination (*i.e.* characterization) then, the performance of the mammogram together with AI was studied in correlation with the combined examination of the mammogram and ultrasound.

To our knowledge, this is a leading study in evaluating mammogram scanned with AI as an alternative tool to the use of the commonly used breast imaging two modalities which are the mammogram and the ultrasound.

The study performed by Dheeba and Selvi showed one of the highest sensitivity of 96.9% and specificity of 92.9% for the proposed AI algorithm in detection of the cancer in the mammograms.[18]

Figure 4. A 45-year-old female presenting with breast density ACR type b and right breast mass proved to be mucinous carcinoma. A, digital mammogram of the right beast (mediolateral oblique and craniocaudal views) right breast lower central (six clock) region showed dense partly irregular mass (arrow). B, breast ultrasound showed upper row grayscale and lower row color-coded; there is a superficial isoechoic purely solid mass, partly irregular that went with the plane of the breast tissue and showed scanty vascularity. The mass was considered probably benign and given BI-RADS 3 (*i.e.* false negative). C, AI heatmap showed the mass with color overlay and a high abnormality scoring of 84%. According to AI the mass is highly suspicious of being malignant and so it was a true-positive case.



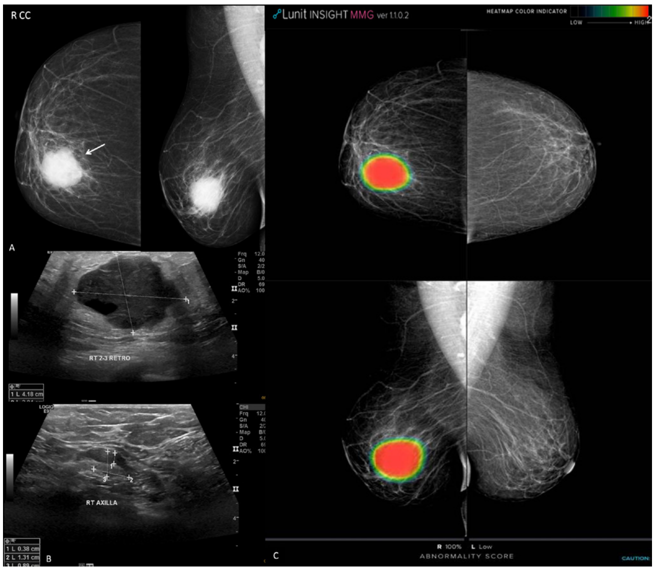
The current work showed sensitivity of 96.8% and specificity of 90.1% for the AI in the discrimination of the breast lesions detected on the included mammograms.

In the current experience; mammograms that were supported with the AI algorithms were able to diagnose 96.8% (n = 622/642) true positive breast lesions, in a way comparable to the combined assessment of the mammogram and ultrasound that showed correct positive diagnoses in 98.6% (n = 633/642), Figure 1. Yet even, the applied color hue provided by the AI scanning presented a precise full extend of the disease in the form of intense hot colored hue (*i.e.* presence of red, yellow and orange colors) not just demarcation of the site of the abnormality, **figures 1c, 1d, 6a and 6b**.

According to Stavros et al, 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.[19]

This was the condition with some cases who presented by suspicious looking lesions on the primary evaluation by mammogram and ultrasound and then when these mammograms were scanned by AI, the suspicious looking lesions were given a faint blue color hue and a low abnormality scoring percentage of suspicion (Figures 2 and 3). In these cases, if the AI pattern of interpretation were considered in the clinical setting, then biopsy could have

Figure 5. Female patient 50 years old with right breast malignant phylloids tumor. A, digital mammogram of the right breast that showed upper inner dense mass was seen framed by a tissue reaction (arrow) that was given a BI-RADS 4 category. B, breast ultrasound (grayscale). Upper row showed a mainly solid isoehoic mass with few cystic changes and marginal lobulations. The mass was seen extending along the axis of the breast tissue and was assigned probably benign (BI-RADS 3). Lower row showed ipsilateral axillary node, with minimal cortical thickening and preserved fatty hilum, and was considered a reactionary node. C, AI scanned mammogram showed intense color hue of mainly red color and a very confident abnormality scoring percentage of malignancy of 100%. The ultrasound downgraded the BI-RADS category of the right breast mass and so the current case was a false negative one by the conventional breast imaging yet the category was upgraded to BI-RADS 5 and the case was true positive on AI basis. AI, artificial intelligence; BI-RADS, Breast Imaging Reporting and Data System

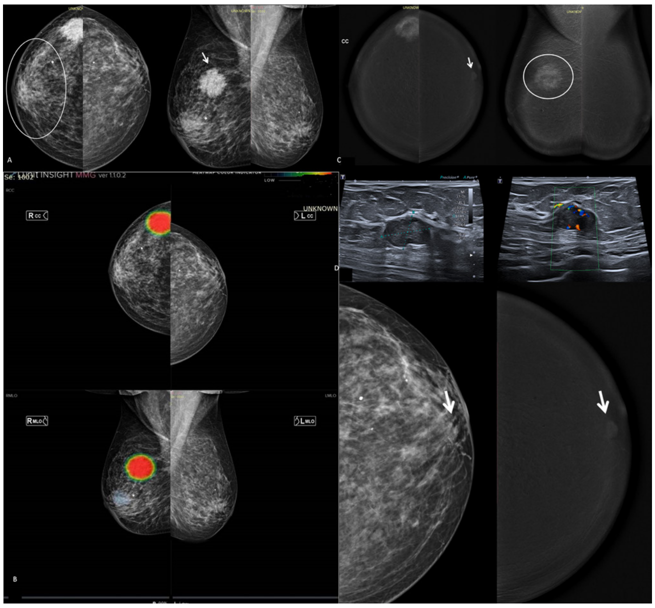


been dismissed or even if biopsy was performed, that would be just to exclude malignancy without unneeded panic and/or anxiety for the patients. These lesions with low scoring percentage will be subjected to interval follow-up and if changes did occur later (whether in the form of increase in the extension marked by the color hue on the mammogram and /or increase in the value of the abnormality scoring percentage that could be visually assessed by the increased in the hotness of the color hue), then at that stage biopsy could be an essential.

AI was able to suggest benign disease and presented true negative cases in 485/538 (90.1%).

On using conventional breast imaging such as mammography and ultrasound, sometimes there is a difficulty to distinguish malignant masses if they possess a benign feature such as a well-circumscribed margin, homogenous texture or oval shape.[20] These types of masses may include: intracystic papillary carcinoma, invasive papillary carcinoma, mucinous, medullary, and metaplastic carcinomas, as well as malignant phyllodes tumors.

Figure 6. Female patient 50 years old with right breast invasive ductal carcinoma and left breast complex adenoma. A, bilateral digital mammogram that showed a right breast upper outer quadrant indistinct dense mass (arrow). Associate areolar and periaerolar mild dermal thickening and the central portion of the breast showed regional area of asymmetrical and increased densities not specified to be part of the disease or just reactionary inflammatory edematous changes. B, AI scanned mammograms that showed intense mainly red color hue applied to the upper outer quadrant mass, yet the central portion of the barest was not included in the heatmap. C, bilateral contrast enhanced mammogram that showed heterogeneous intense contrast uptake of only the upper outer quadrant right breast mass (circle) and so contrast-enhanced mammogram confirmed the scanning presented by the AI and so accordingly, the case was diagnosed right breast unifocal carcinoma and associate reactive inflammatory changes. The left breast showed a small homogenously enhancing circumscribed mass (arrow) that was not spotted on the AI scanned images. D, breast ultrasound, left image is a grayscale and right image is a color-coded mode, showed the left breast retroareolar solid mass of increased vascularity that proved to be complex adenoma. E, crainocoded view of the left breast, left image is a digital mammogram and right image is a contrast-enhanced mammogram, retrograde evaluation of the conventional mammogram guided by the contrast enhanced image; the left breast complex adenoma_ that was ignored by the AI_ was located on the mammogram (arrow). AI, artificial intelligence.



In the current work, the combined consideration of the color hue for lesion localization and the value of the abnormality scoring presented by the AI helped to diagnose carcinoma in probably benign looking masses (Figures 4 and 5).

For malignant looking and suspicious lesions that were assigned by the readers as BI-RADS 4 and 5 category and were proved to be cancer by the pathology ($n = 623/642$) represented 97% with a range of abnormality scoring at the AI of 59 to 100%.

On the other side, breast lesions with typical benign or probably benign breast features on the conventional imaging that were assigned BI-RADS 2 and 3 categories respectively and proved to be benign by the pathology (n = 404 lesions, 75%) showed a range from zero (i.e. >10%) to 40% on AI scanning.

Pathology proved benign lesions showed AI median value of the abnormality scoring percentage of 25 and for malignant lesions, it was 88.

AI abnormality scoring of 59% yielded a specificity of 90.1%, NPV of 96% and accuracy of 93.8% in the discrimination of the breast lesions that were detected on the included mammograms.

False-negative lesions on AI (n = 20) were invasive ductal carcinoma (n = 4), invasive lobular (n = 4), DCIS (n = 6), inflammatory carcinomatosis (n = 5), and medulllary carcinoma (n = 1). While, false-positive lesions (n = 53) were fibroadenolipoma (n = 5), fibrocystic disease (n = 13), fat necrosis (n = 7), fibroadenoma (n = 3), infectious mastitis (n = 15), fibrous mastopathy (n = 3) and granulomatous mastitis (n = 7),

In the current experience; the AI algorithm sometimes ignored typical benign looking masses. As a result, no focal areas were spotted by the color hue at the heatmap images, the given abnormality scoring for the breast as whole was a low scoring of <10% and the breast was assigned as negative for malignancy. However, some of these mammograms showed obvious benign looking masses or masses that were only conspicuous on the ultrasound.

Such final outcome is not accepted in the clinical practice as the presence of benign masses requires close follow-up not neglection and part of them may even require elective biopsy/ surgery Figure 6.

So, in case of dense breast (*i.e.* ACR c and d), it is recommended to support the AI-mammogram findings with an ultrasound examination.

Another point that could limit the inclusion of the AI systems in the clinical setting is the inability of the machine to explain the presented data to the human user.

Human-like AI systems learn from both examination-level and pixel-level labels and require a detailed supervision to be able to supply a high performance. However, inter observer variability does widen the scope.[21]

Medical images have properties that make them very different from images from natural scenes (*e.g.* images of a tree or dog), *e.g.* the objects of interest that determine the class usually occupy a large fraction of natural images, objects of interest on medical images are often relatively small, which sometimes make it difficult for AI system to analyze unlike natural scenes (*e.g.* images of a plant or an animal) where objects of interest usually occupy a large fraction of these natural images.[3]

In the current experience, there was a significant correlation between the BI-RADS category that was assigned by the combined evaluation of the mammogram and ultrasound 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), “likely malignant, BI-RADS 5” (p value:<0.001).

The radiologists need many years to learn from millions of images, yet a neural network can learn and modulate in a few days. Therefore, it is expected that human-like AI machines are going to be used as a creditable method of knowledge discovery so long as their performance is adequately supervised and consequently their ability to explain predictions is enhanced.

Further research studies are needed to pave the way for these neural networks and indulge them for optimization of the mammograms in particular and generally medical images.

CONCLUSION

AI could be considered as an optional primary reliable complementary tool to the digital mammogram for the evaluation of the breast lesions. The color hue and the abnormality scoring percentage presented a credible method for the localization and discrimination of breast cancer of near accuracy to the breast ultrasound. So consequently, AI- mammogram combination could be used as a one setting method to discriminate between breast lesions that require further imaging or biopsy from those that need only time interval follow-up.

AUTHORS’ CONTRIBUTIONS

1. Guarantor of integrity of the entire study: M.S.
2. Study concepts and design: K.R. and M.S.
3. Literature research: M.S., H.L., and A.B.
4. Clinical studies: H.L. and A.B.
5. Experimental studies / data analysis: M.S., H.L., and A.B.
6. Statistical analysis: M.S.
7. Manuscript preparation: K.R., H.L., A.B. and M.S..
8. Manuscript editing: K.R., and M.S.

CONSENT FOR PUBLICATION

All patients included in this research were legible; above 16 years of age. They gave written informed consent to publish the data contained within this study.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the ethical committee of the Radiology Departments of an academic highly specialized multidisciplinary Hospital and an informed written consent was taken from the patients that were included in the study.

REFERENCES

1. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 2002; **225**: 165–75. doi: https://doi.org/10.1148/radiol.2251011667
2. Rosenberg RD, Hunt WC, Williamson MR, Gilliland FD, Wiest PW, Kelsey CA, et al. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, new Mexico. *Radiology* 1998; **209**: 511–8. doi: https://doi.org/10.1148/radiology.209.2.9807581
3. Geras KJ, Mann RM, Moy L. Artificial intelligence for mammography and digital breast Tomosynthesis: current concepts and future perspectives. *Radiology* 2019; **293**: 246–59. doi: https://doi.org/10.1148/radiol.2019182627
4. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. *Beth Israel Deaconess Medical Center, Harvard Medical School* 2016;: 1. −−6p..
5. Bahl M. Detecting breast cancers with mammography: will AI succeed where traditional CAD failed? *Radiology* 2019; **290**: 315–6. doi: https://doi.org/10.1148/radiol.2018182404
6. Nam H, Kim H-E. Batch-instance normalization for adaptively style-invariant neural networks. *Adv Neur In* 2018; **1**: 2563–72.
7. Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. *Proc IEEE I Conf Comp Vis* 2015; **1**: 1520–8.
8. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA, et al. ACR BI-RADS® atlas, breast imaging reporting and data system. *Reston, VA, American College of Radiology* 2013;.
9. Sadoughi F, Kazemy Z, Hamedan F, Owji L, Rahmanikatigari M, Talebi Azadboni T, Azadboni T. Artificial intelligence methods for the diagnosis of breast cancer by image processing: a review. *Breast Cancer - Targets and Therapy* 2018; **10**: 219–30. doi: https://doi.org/10.2147/BCTT.S175311
10. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography compared to mammography alone in women at elevated risk of breast cancer: results of the first-year screen in ACRIN 6666. *JAMA* 2008; **299**: 2151–63.
11. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, et al. Canadian association of radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J* 2018; **69**: 120–35. doi: https://doi.org/10.1016/j.carj.2018.02.002
12. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017; **2**: 230–43. doi: https://doi.org/10.1136/svn-2017-000101
13. Shen D, Wu G, Zhang D, Suzuki K, Wang F, Yan P. Machine learning in medical imaging. *Comput Med Imaging Graph* 2015; **41**: 1–2. doi: https://doi.org/10.1016/j.compmedimag.2015.02.001
14. Allen B, Dreyer K. The artificial intelligence ecosystem for the radiological sciences: ideas to clinical practice. *J Am Coll Radiol* 2018; **15**: 1455–7. doi: https://doi.org/10.1016/j.jacr.2018.02.032
15. Collier M, Fu R, Yin L. Artificial intelligence: healthcare's new nervous system.Published 2017. Available from: Accenture website. www.accenture.com/us-en/insight-artificialintelligence-healthcare [Accessed February 2, 2021].
16. Stavros AT, Freitas AG, deMello GGN, Barke L, McDonald D, Kaske T, et al. Ultrasound positive predictive values by BI-RADS categories 3–5 for solid masses: an independent reader study. *Eur Radiol* 2017; **27**: 4307–15. doi: https://doi.org/10.1007/s00330-017-4835-7
17. Mendelson EB. Artificial intelligence in breast imaging: potentials and limitations. *AJR Am J Roentgenol* 2019; **212**: 293–9. doi: https://doi.org/10.2214/AJR.18.20532
18. Dheeba J, Tamil Selvi S, Selvi TS. An improved decision support system for detection of lesions in mammograms using differential evolution optimized wavelet neural network. *J Med Syst* 2012; **36**: 3223–32. doi: https://doi.org/10.1007/s10916-011-9813-z
19. Stavros AT, Freitas AG, deMello GGN, Barke L, McDonald D, Kaske T, et al. Ultrasound positive predictive values by BI-RADS categories 3-5 for solid masses: an independent reader study. *Eur Radiol* 2017; **27**: 4307–15. doi: https://doi.org/10.1007/s00330-017-4835-7
20. Yoo JL, Woo OH, Kim YK, Cho KR, Yong HS, Seo BK, et al. Can MR imaging contribute in characterizing well-circumscribed breast carcinomas? *Radiographics* 2010; **30**: 1689–704. doi: https://doi.org/10.1148/rg.306105511
21. Buelow T, Heese HS, Grewer R, Kutra D, Wiemker R. Inter- and intra-observer variations in the delineation of lesions in mammograms. medical imaging 2015: image perception, observer performance, and Technology assessment. *Bellingham, Wash: International Society for Optics and Photonics* 2019; **941605**.