

UC Irvine

UC Irvine Previously Published Works

Title

Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning With Consideration of Peritumor Tissue

Permalink

<https://escholarship.org/uc/item/1jw642gg>

Journal

Journal of Magnetic Resonance Imaging, 51(3)

ISSN

1053-1807

Authors

Zhou, Jiejie
Zhang, Yang
Chang, Kai-Ting
[et al.](#)

Publication Date

2020-03-01

DOI

10.1002/jmri.26981

Peer reviewed



Published in final edited form as:

J Magn Reson Imaging. 2020 March ; 51(3): 798–809. doi:10.1002/jmri.26981.

Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning with Consideration of Peri-Tumor Tissue

Jiejie Zhou, MD¹, Yang Zhang, PhD², Kai-Ting Chang, MD², Kyoung Eun Lee, MD, PhD³, Ouchen Wang, MD⁴, Jiance Li, MD¹, Yezhi Lin, PhD⁵, Zhifang Pan, PhD⁵, Peter Chang, MD², Daniel Chow, MD², Meihao Wang, MD, PhD^{1,*}, Min-Ying Su, PhD^{2,**}

¹Department of Radiology, The First Affiliate Hospital of Wenzhou Medical University, Wenzhou, China

²Department of Radiological Sciences, University of California, Irvine, CA, USA

³Department of Radiology, Inje University Seoul Paik Hospital, Inje University, Seoul, Korea.

⁴Department of Thyroid and Breast Surgery, The First Affiliate Hospital of Wenzhou Medical University, Wenzhou, China

⁵Information Technology Center, The First Affiliate Hospital of Wenzhou Medical University, Wenzhou, China

Abstract

BACKGROUND: Computer-aided methods have been widely applied to diagnose lesions detected on breast MRI, but fully-automatic diagnosis using deep learning is rarely reported.

PURPOSE: To evaluate the diagnostic accuracy of mass lesions using ROI-based, radiomics and deep learning methods, by taking peri-tumor tissues into consideration.

STUDY TYPE: Retrospective

POPULATION: 133 patients with histologically confirmed 91 malignant and 62 benign mass lesions for training (74 patients with 48 malignant and 26 benign lesions for testing).

FIELD STRENGTH/SEQUENCE: 3T, using the volume imaging for breast assessment (VIBRANT) DCE sequence.

ASSESSMENT: 3D tumor segmentation was done automatically by using fuzzy-C-means algorithm with connected-component labeling. A total of 99 texture and histogram parameters were calculated for each case, and 15 were selected using random forest to build a radiomics model. Deep learning was implemented using ResNet50, evaluated with 10-fold cross-validation. The tumor alone, smallest bounding box, and 1.2, 1.5, 2.0 times enlarged boxes were used as inputs.

*Corresponding Author: Meihao Wang, MD (in China), Department of Radiology, The First Affiliate Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, P.R. China. Tel: +86 13806683327, Fax: +86 0577 86699033, wzwmh@wmu.edu.cn.

**Corresponding Author: Min-Ying Su, PhD (in United States), 164 Irvine Hall, Center for Functional Onco-Imaging, University of California, Irvine, CA 92697-5020. Tel: +1 (949) 824-4925, Fax: +1 (949) 824-3481, msu@uci.edu.

STATISTICAL TESTS: The malignancy probability was calculated using each model, and the threshold of 0.5 was used to make diagnosis.

RESULTS: In the training dataset, the diagnostic accuracy was 76% using three ROI-based parameters, 84% using the radiomics model, and 86% using ROI+radiomics model. In deep learning using per-slice basis, the area under the ROC was comparable for tumor alone, smallest and 1.2 times box (AUC=0.97–0.99), which were significantly higher than 1.5 and 2.0 times box (AUC= 0.86 and 0.71, respectively). For per-lesion diagnosis, the highest accuracy of 91% was achieved when using the smallest bounding box, and that decreased to 84% for tumor alone and 1.2 times box, and further to 73% for 1.5 times box and 69% for 2.0 times box. In the independent testing dataset, the per-lesion diagnostic accuracy was also the highest when using the smallest bounding box, 89%.

DATA CONCLUSION: Deep learning using ResNet50 achieved a high diagnostic accuracy. Using the smallest bounding box containing proximal peri-tumor tissue as input had higher accuracy compared to using tumor alone or larger boxes.

Keywords

Breast cancer diagnosis; DCE-MRI; Deep learning; Peri-tumor tissue; Radiomics; ResNet

Breast MRI is an important imaging modality for screening, diagnosis and pre-operative staging of breast cancer.^{1,2} Many benign lesions also show strong contrast enhancement, and may lead to false positive diagnosis, unnecessary biopsy or over treatment. With increasing screening and preoperative MRI performed, particularly in community settings,³ an efficient way for characterization of the enhancing lesions is important to improve diagnostic accuracy.

Conventional diagnosis made by radiologists is mainly based on evaluation of the morphological features and the DCE time course, which is subjective and varies with radiologists' experience. This problem was well recognized, and many computer-aided-diagnosis (CAD) methods have been developed and reported in the literature in the last two decades.^{4–8} In addition to providing quantitative parameters related to shape, internal heterogeneity and DCE kinetics, the CAD features were further related to BI-RADS descriptors,^{5,6} and used to build separate diagnostic models for mass and non-mass-like enhancements, respectively.^{7,8} With the advances in computer technology, extracting large data from medical images using automatic algorithms becomes more feasible; and “radiomics”, which allows high-throughput extraction of tremendous amount of quantitative information from radiographic images, emerged.^{9,10} Texture and histogram features based on MR images have potential to provide noninvasive imaging biomarkers to aid in breast cancer diagnosis, prognosis and treatment response evaluation.^{11,12} The radiomics signatures are also related to molecular biomarkers and subtypes, and can aid in patients' management using precision medicine approach.^{13,14}

In recent years, artificial intelligence (AI) algorithms, particularly deep learning, have demonstrated remarkable progress in medical image analysis, advancing the field forward at a rapid pace.¹⁵ Convolutional Neural Network (CNN) is a common deep learning method

applied to analyze photographic, pathological and radiographic images, and reported to have great potential in various clinical tasks such as segmentation, abnormality detection, disease classification and diagnosis.¹⁶ Deep learning has been applied to detect and diagnose breast cancer on mammography, and shows promising results. For mass lesions, the accuracy of deep learning was comparable to that of experienced radiologists.^{17–22} Breast MRI acquires multiple sets of images with varying tissue contrast, and DCE-MRI further acquires images at different times with varying signal intensities that need to be considered, which makes implementation of deep learning algorithms more challenging, and rarely reported.^{23–25} Truhn et al. investigated the diagnostic performance of benign and malignant lesions in MRI using radiomics and deep learning.²⁵ In their study, the input box was much larger than the size of small lesions, which contained the suspicious lesion with a large amount of peri-tumor and normal tissues, and might affect the diagnostic performance.

The tumor microenvironment is known to play a very important role in growth and invasion of tumor,^{26,27} and peri-tumor tissue has been shown to provide helpful information for diagnosis and prediction of prognosis.^{28–32} However, how the peri-tumor tissue should be evaluated has not been well studied.³¹ The main goal of this study is to evaluate the diagnostic accuracy of breast lesions detected on DCE-MRI with deep learning, by using 5 different sizes of input boxes containing the tumor with different amount of peri-tumor tissues to evaluate their diagnostic performance. For comparison with the deep learning results, the diagnosis was also done with conventional methods using the whole tumor ROI-based analysis (tumor size, volume and enhancement ratios) and radiomics.

MATERIALS AND METHODS

Patients

This is a retrospective study. A total of 133 patients were used in the training dataset, including 91 malignant cancers (mean age 51 ± 10), and 62 benign lesions (mean age 45 ± 11). All lesions were confirmed by histological examination, major types listed in Table 1. These cases were selected from consecutive patients receiving breast MRI for diagnosis from January 2017 to May 2018, before biopsy or any treatment. All studies with confirmed pathological diagnosis were selected. Since one major purpose of this study was to evaluate the peri-tumor tissues surrounding the lesion, a well-defined tumor boundary was needed, and thus only mass lesions that were visible on contrast-enhanced images were included. Also, to ensure peri-tumor tissue was present for analysis, large tumors with volume greater than 12 cm^3 were excluded. For independent testing, the newer cases performed from June to Dec 2018 were used, based on the same selection criteria. This study was approved by the ethics committee of our hospital, and informed consent was waived.

MRI Protocol and Tumor Segmentation

All patients underwent MRI on a 3T scanner (GE SIGNA HDx) using a dedicated 8-channel bilateral breast coil. The dynamic-contrast-enhanced (DCE) scan was acquired using the volume imaging for breast assessment (VIBRANT) sequence in the axial view to cover both breasts, with TR=5 ms; TE=2 ms; FA=10°; slice thickness=1.2 mm; FOV=34×34cm²; matrix size=416×416. The DCE series consisted of 6 frames: one pre-contrast (F1) and 5

post-contrast (F2-F6). The acquisition time for each frame was 1 min 32 s. The contrast agent, 0.1 mmol/kg gadopantetate dimeglumine (Magnevist; Bayer Schering Pharma), was intravenously injected after the pre-contrast images were acquired, at a rate of 2 ml/s followed by 20 ml saline flush at the same rate.

A radiologist reviewed the images, and indicated the location and the slice range that contained the tumor, by referencing to the clinical, radiological and pathological reports. Then, based on the information, the tumor ROI on all slices were automatically segmented on contrast-enhanced maps by using the fuzzy-C-means (FCM) clustering algorithm with 3D connected-component labeling, as described previously.^{5,7,33} The lesion location and range information was provided to another radiologist to perform segmentation again, and the obtained radiomics features were compared to test their reproducibility by using intra-class-coefficient (ICC).

ROI-based and Radiomics Analysis

Three heuristic DCE parametric maps were generated according to:

$$\text{Wash-in Signal Enhancement (SE) Map} = [(F2-F1) / F1]$$

$$\text{Maximum Signal Enhancement (SE) Map} = [(F3-F1) / F1]$$

$$\text{Wash-out Slope Map} = [(F6 - F3) / F3]$$

The generated DCE parametric maps were inspected to make sure no motion artifact. Figure 1 shows the DCE images (F1, F2, F3 and F6), segmented tumor, three parametric maps, and the mean DCE time course of a benign fibroadenoma. Figure 2 shows the images of an invasive ductal carcinoma.

On each parametric map, 20 Gray Level Co-occurrence Matrix (GLCM) texture features,³⁴ and 13 histogram-based parameters (10%, 20%... 80% to 90% values, mean, standard deviation, kurtosis and skewness) were calculated, with a total of 99 quantitative pixel-wised imaging features. The tumor segmentation was done on each 2-D slice, and they were rendered into a 3-D space with isotropic voxel resolution for extracting the 3D texture features. The intra-class-coefficient (ICC) of features analyzed between the two radiologists was 0.91 ± 0.11 , showing a high reproducibility. This was likely due to the analysis of only mass lesions in this study, and also that the segmentation was done using a computer-program, not manually. Therefore, the reproducibility of the extracted radiomics features was not used as a pre-selection criterion for feature reduction.

After the features were extracted for all cases, they were properly normalized to mean=0 and standard deviation=1. The random forest algorithm with bootstrap-aggregated decision trees was applied to select features to build an optimal diagnostic model.³⁵ The first step was to select important ones and rank the discriminating significance of all features, by using a total of 1,000 trees. During the permutation process, each feature and case could be extracted hundreds of times. The curvature test was implemented during the process of parameter tuning to select uncorrelated features. The significance of each feature was determined based on the decrease of classification accuracy when this feature was removed. The diagnostic performance was tested using 10-fold cross-validation, which could avoid over-fitting and

also improve the general applicability of the developed model. The final diagnostic model was built by logistic regression, first by using the top 20 features, and then by removing the lowest one, two, three ... one by one. The AUC started to show substantial decrease after removing 5 features; therefore, the final model was built with 15 features. The detailed radiomics analysis and model-building procedures were described in a recent publication.³⁶ The analysis was done using programs written in Matlab 2013b (The Mathworks Inc.).

Five whole tumor ROI-based parameters, including the 1-D tumor size, 3-D tumor volume, mean Wash-in SE ratio, mean Maximum SE ratio and mean Wash-out slope, were calculated. The mean values in the malignant and benign groups of the training and testing datasets are shown in Table 2. Three ROI-based parameters that gave the best classification performance were selected to train a logistic model for diagnosis. Then, these three ROI-based parameters and 15 radiomics features were used to build a combined ROI+radiomics model.

Deep Learning Analysis

Deep learning was applied to automatically differentiate the two groups, by using ResNet50 architecture. The conventional convolutional neural network (CNN) learns features using large convolutional network architectures; and in contrast, the ResNet tries to extract residual features, as subtraction of features learned from input of that layer, using “skip connections”.³⁷ The ResNet50 architecture contains one 3×3 convolutional layer, one max pooling layer, and 16 residual blocks. Each block contains one 1×1 convolutional layer, one 3×3 convolutional layer and one 1×1 convolutional layer. The residual connection is from the beginning of the block to the end of the block. The output of the last block was connected to a fully-connected layer with sigmoid function to give the prediction. The methods were similar to those used in Haarbuerger et al.^{24,25} The software code was written in Python 3.5 using the open-source TensorFlow r1.0 library (Apache 2.0 license), on a GPU-optimized workstation with a single NVIDIA GeForce GTX Titan X (12GB, Maxwell architecture).

The analysis was done by using three DCE parametric maps as inputs. For each case, the smallest square bounding box containing the entire tumor was generated. This was done by projecting the segmented tumor ROI's from all slices together, and the smallest square box covering the projected boundary was generated.³⁸ In order to evaluate the diagnostic role of peri-tumor tissues, 5 different input boxes were used, including 1) the tumor alone by setting all outside tumor pixels in the box as zero, 2) the smallest bounding box, 3) enlarged box by 1.2 times, 4) enlarged box by 1.5 times, and 5) enlarged box by 2.0 times. The same box was used for all slices in one case. The input boxes of two benign cases are illustrated in Figure 3, and those of two malignant cases are shown in Figure 4.

The bounding box was resized to 75×75 pixels as input into the networks. All tumor slices were used as independent inputs, and the dataset was further augmented 20 times by using random affine transformations. The loss function was cross entropy. The training was implemented using the Adam optimizer fixed to 0.001.³⁹ Parameters were initialized using ImageNet.⁴⁰ The L2 regularization was performed to prevent over-fitting of data by limiting the squared magnitude of the kernel weights. Additionally, an early stopping strategy was

used, in which the same epoch number was applied to all folds in cross validation. The classification performance was evaluated using 10-fold cross-validation, and each case had only one chance to be included in the validation group. According to the predicted malignancy probability for each slice, the results from all slices were combined to generate the ROC curve.

The prediction results based on 2D slices meant each slice had its own diagnostic probability. For per-lesion diagnosis, the highest probability among all slices of one lesion was considered. Using this definition could increase the false positive rate, and to investigate this, the results obtained using different threshold values were compared.

Statistical Analysis

The statistical analysis was performed with SPSS 16.0, with $P < 0.05$ considered significant. In ROI-based, radiomics and ROI+radiomics analysis, after the model was built, the malignancy probability for each lesion was calculated, and they were used for ROC analysis. In addition, a diagnosis was made for each lesion, based on the threshold of probability 0.5 as malignant, and then the sensitivity, specificity and accuracy were calculated. In deep learning, the analysis was done using each slice as an individual input, and the obtained malignancy probability from all slices were combined for performing the ROC analysis. The curves obtained using 5 different input boxes were compared by using the DeLong test, with $\alpha = 0.05$. The respective models developed in the training dataset were applied to the independent testing dataset to give a diagnosis for each lesion, and then the sensitivity, specificity and accuracy were calculated.

RESULTS

ROI-based Volume and Mean DCE Parameters

Three parameters, including 3D tumor volume, wash-in SE ratio and wash-out slope, that gave the best classification performance were combined to train a logistic model, and the overall diagnostic accuracy was 76%. The diagnostic sensitivity, specificity and accuracy are summarized in Table 3. The model developed from the training dataset was applied to the testing dataset, and the accuracy was 67%.

Radiomics Analysis

The results are shown in Table 3. The 15 selected radiomics features and the diagnostic model built by logistic regression are included in the supplementary material. The plot of the malignancy probability based on the final radiomics model is shown in Figure 5. The diagnostic accuracy was 84%. When combining the three whole tumor ROI-based parameters and the 15 selected radiomics features together, the accuracy was improved to 86%. The combined diagnostic model is also included in the supplementary material. When applying these models to the testing dataset, the accuracy was 78% for radiomics, and 77% for ROI+radiomics.

Deep Learning Analysis Using ResNet50

The results obtained using 5 different input boxes containing different amount of peri-tumor tissues were compared. The mean 3D tumor volumetric percentage in the smallest bounding box was 34%. In ROC analysis performed using the predicted per-slice malignancy probability, the AUC was 0.97 ± 0.03 (range 0.93–0.99) for tumor alone, 0.98 ± 0.03 (range 0.90–0.99) for smallest bounding box, 0.99 ± 0.01 (range 0.97–0.99) for 1.2 times box, 0.86 ± 0.07 (range 0.76–0.92) for 1.5 times box, 0.71 ± 0.06 (range 0.63–0.81) for 2.0 times box. The ROC curves obtained using these 5 different input boxes are shown in Figure 6. The DeLong test showed that the ROC of smallest bounding box was not different compared to tumor alone ($z=1.37$, $p=0.42$) and 1.2 times box ($z=1.15$, $p=0.13$), and significantly better than 1.5 times ($z=2.74$, $p=0.01$) and 2.0 times boxes ($z=3.25$, $p<0.00001$).

According to the per-slice results, the highest probability in one lesion was used to make per-lesion diagnosis, using the threshold of 0.5. The results are also shown in Table 3. When using the tumor alone, the sensitivity was $91/91=100\%$, the specificity was $38/62=61\%$, with the overall accuracy of 84%. When using the smallest bounding box, the sensitivity was $90/91=99\%$, the specificity was $49/62=79\%$, with the overall accuracy of 91%. The results showed that when considering adjacent peri-tumor using the smallest bounding box compared to using tumor alone, the false positive case was decreased from 24/62 to 13/62, and that improved the specificity from 61% to 79% and the accuracy from 84% to 91%. When using the enlarged boxes with more peri-tumor tissue, the prediction accuracy became worse and worse as the box became bigger and bigger. The results in the testing dataset showed similar trends. The accuracy for the per-lesion diagnosis was 89% when using the smallest bounding box, and worse for larger boxes.

Per-lesion Diagnosis Based on Different Malignancy Probability Threshold

The diagnostic results of the 4 illustrated case examples are given in the figure legends. For the imaging slice shown in Figure 1, the predicted malignancy probability using ResNet was 0.36, correctly diagnosed as benign. However, for the whole lesion, one out of the total of 14 slices had the highest malignancy probability of 0.51, leading to a wrong malignant diagnosis according to the threshold of 0.5. If the threshold was set higher, this case could be correctly diagnosed. In order to investigate the trade-off between sensitivity and specificity, the results obtained with varying threshold from 0.5 to 0.7 were compared, listed in Table 4. As expected, increasing the threshold value could improve the specificity, with decreased sensitivity. By using the threshold of 0.5, 0.55, 0.6, 0.65 and 0.7 in the testing dataset, the specificity was 81%, 81%, 92%, 92% and 100%, with accuracy of 89%, 89%, 81%, 78% and 53%, respectively.

DISCUSSION

In this study we evaluated the diagnostic performance of breast mass lesions detected on DCE-MRI using ROI-based, radiomics, and deep learning methods with different input box sizes. In the training dataset, the accuracy was 76% using ROI-based parameters, 84% using radiomics, and 86% using combined ROI+radiomics. In deep learning using ResNet50 with the smallest bounding box as input, the accuracy was improved to 91%. The results obtained

in the testing dataset using newer cases were comparable, showing exactly the same trend with a slightly lower accuracy. The results suggest that deep learning has the potential to be developed as a clinical diagnostic tool.

In deep learning, the selection of the input box has never been systematically studied before, and thus, one purpose of this work was to perform and compare the results done using 5 different methods, by considering different amount of peri-tumor tissues. Previous studies have shown that the peri-tumor environments contain important information related to the aggressiveness of the tumor, reflecting lymphovascular invasion and angiogenesis,^{28,29} composition of lipid and edema,^{30–32} or mammary field cancerization,^{41,42} and that can be used for prediction of diagnosis or prognosis. In this study we used different sizes of bounding box as inputs to evaluate their diagnostic role. In per-lesion diagnosis, the accuracy was the highest when using the smallest bounding box. As the size of the box increased, the performance became worse and worse, which might be due to the diluted information by containing too much normal tissue, as well as the degraded input image resolution into the neural networks. We further investigated the trade-off between sensitivity and specificity using the results of the smallest bounding box. If the lesions are enhanced and already determined as abnormal that needs further attention, the threshold can be set differently depending on the intended clinical applications. For example, if the goal is to rule out malignancy, the threshold can be set higher than 0.5 to decrease false positive diagnosis.

The role of peri-tumor tissue at various distances away from the tumor in predicting tumor aggressiveness has been investigated before. Shin et al. applied a shell-based method and reported that the apparent diffusion coefficient (ADC) of proximal peritumoral stroma could differentiate between low-risk and high-risk breast cancer, but not the middle or the distal peritumoral stroma.³¹ Fan et al. also applied a similar method and found proximal peritumoral stroma could differentiate between low and high Ki-67 breast cancer groups.⁴³ The tissues further away from the tumor boundary contained less information associated with the tumor, thus could be interpreted as “normal”; however, there was no definition of the cut-off distance that could be used to classify tissues into “peri-tumor” vs. “normal”. Our results also agreed that by taking the proximal peri-tumor tissue into consideration, i.e. by using the smallest bounding box as input in deep learning, it could achieve a higher diagnostic accuracy compared to using tumor alone or larger boxes.

In addition to deep learning, we also performed diagnosis using traditional tumor ROI-based model and the more sophisticated radiomics model for comparison. Since malignant tumors were more likely to be bigger and showing the wash-out DCE pattern with stronger enhancements, using a simple ROI-based model could achieve a decent accuracy, 76% in the training dataset. Radiomics could evaluate the internal heterogeneity by using texture and histogram analysis, and the accuracy was improved to 84%, with 14 of 15 selected features from texture. Our accuracy was comparable to that Truhn et al., who reported the AUC of 0.78–0.81 for radiomics.²⁵ In another study by Whitney et al. to differentiate between benign and Luminal A breast cancer, the AUC was 0.68 using maximum linear size, and 0.73 using radiomics features.⁴⁴ Since the radiomics features were extracted from the segmented or manually contoured tumor according to the precise boundary, the margin might not be well evaluated. Kooi et al.¹⁸ used an expanded area to compute the margin

contrast on mammography, which may be implemented on MRI to evaluate whether it can improve the diagnostic accuracy of detected lesions using radiomics.

In our deep learning, ResNet50 was used as the architecture of the convolutional neural network. Deep learning with various CNN architecture has been applied to differentiate benign and malignant mass lesions on mammography.^{17–22} Chougrad et al. used three different CNN, and reported that ResNet50 could reach convergence during optimization process faster than VGG, and obtain a good accuracy.²⁰ Our ResNet50 method was similar to ResNet18 and ResNet34 used in Haarburger et al. and Truhn et al.^{24,25} In our study, each slice was used as individual input, and L2 norm regularization, dropout and data augmentation were applied to control overfitting. In per-slice analysis using 10-fold cross-validation, the AUC's were > 0.90 in all runs, suggesting that the trained model was robust and not over-fitted. In ResNet, since it was pre-trained with photographs with RGB colors, only 3 sets of images can be used in input channel. Haarburger et al. investigated various combinations and found that the pre-contrast F1, post-contrast F3 and subtraction (F2-F1) gave the best accuracy.²⁴ In the present study we used three generated DCE parametric maps as inputs, $(F2-F1)/F1$ and $(F3-F1)/F1$, with $(F6-F3)/F3$ to take the DCE wash-out pattern into account. As T2-weighted images also provide very helpful diagnostic information, other CNN architecture that can consider more sets of images can be investigated in the future.

Two other studies also investigated the application of deep learning for cancer diagnosis on breast MRI. In an earlier study, Antropova et al.²³ used three images as input. In a newer study,⁴⁵ they trained a long short-term memory (LSTM) network which could consider the entire temporal sequences acquired in DCE-MRI, and achieved a significantly improved AUC to 0.88 to differentiate benign from malignant lesions. In their study, the ROI was selected to cover the segmented lesion, similar to our smallest bounding box. Another paper by Zhou et al.⁴⁶ applied weakly supervised 3D deep learning, by using the entire segmented breast as input to predict the presence of benign vs. malignant lesions inside, and obtained AUC of 0.859. However, the main novelty in that study was to localize the lesion, not to diagnose detected lesions. Two review papers by Reig et al.⁴⁷ and Sheth et al.⁴⁸ gave comprehensive information and new research direction about the application of AI and machine learning for analysis of breast MRI.

This study has several limitations. First, the dataset was small for deep learning. For medical image analysis using deep learning, it was usually done by using each slice as an independent input, and the dataset was further enhanced with augmentation; and lastly, appropriate methods such as L2 norm regularization and dropout were used to avoid overfitting. Since this was a training process, the CNN results were usually compared to the conventional ROI-based and radiomics results to evaluate their performances on the same dataset. In the present study, we further used a separate dataset for independent testing, and all results suggest that deep learning can achieve a high accuracy, and has the potential for clinical implementation. Second, for per-lesion diagnosis, the highest malignancy probability among all slices of one lesion was assigned to that lesion. Although this could lead to a high sensitivity, it was at the expense of decreased specificity. The threshold used for diagnosis can be adjusted depending on the intended clinical purpose. Also, how to incorporate the predicted per-slice probabilities from all slices with an optimal weighting to

yield the per-lesion probability needs to be further investigated. Third, in order to investigate the impact of peri-tumor tissue, we only included mass lesions that had a clear boundary in this study. It is known that diagnosis of mass lesions is easier and can achieve a higher accuracy compared to non-mass-like (NML) enhancements. For NML, the tumorous tissues and stroma are mixed, and thus, it is difficult to define the boundary for investigating the role of peri-tumor. Since a clean dataset with well-enhanced mass lesion is used in this study, the developed diagnostic models may not be directly applicable to other datasets. Nonetheless, the models developed in deep learning may provide a basis to be applied to other datasets through proper transfer learning, which is an efficient strategy commonly used in clinical implementation of AI-based diagnostic tools.

In conclusion, we applied ROI-based, radiomics, and deep learning methods to diagnose mass lesions detected on MRI. The results obtained using 5 different input boxes in deep learning, by considering different amount of peri-tumor tissues, were compared. It was shown that deep learning could achieve a higher diagnostic accuracy compared to ROI-based and radiomics models to differentiate benign from malignant lesions. The results also showed that using the smallest bounding box that included small amount of peri-tumor tissue adjacent to the tumor had a higher accuracy compared to using tumor alone or larger input boxes. As many breast MRI is performed in the community settings, the AI-based diagnostic tools may be very helpful. Automatic, computer-aided, diagnosis using artificial intelligence is emerging, and our study may contribute in development of such diagnostic tools in the near future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work was supported in part by Foundation of Wenzhou Science & Technology Bureau (No. Y20180187 and Y20180144), Medical Health Science and Technology Project of Zhejiang Province Health Commission (No. 2019KY102), Zhejiang Provincial Natural Science Foundation of China (No. LQ15A010009 and LY16F030010), and NIH/NCI R01 CA127927 and R21 CA208938.

References:

1. Marino MA, Helbich T, Baltzer P, Pinker-Domenig K. Multiparametric MRI of the breast: A review. *J Magn Reson Imaging* 2018;47:301–315. [PubMed: 28639300]
2. Mann RM, Kuhl CK, Moy L. Contrast-enhanced MRI for breast cancer screening. *J Magn Reson Imaging*. 2019. doi: 10.1002/jmri.26654. [Epub ahead of print]
3. Hill DA, Haas JS, Wellman R, et al. Utilization of breast cancer screening with magnetic resonance imaging in community practice. *J Gen Intern Med* 2018; 33:275–283. [PubMed: 29214373]
4. Gilhuijs KG, Giger ML, Bick U. Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Med Phys*. 1998;25(9):1647–1654. [PubMed: 9775369]
5. Nie K, Chen JH, Yu HJ, Chu Y, Nalcioglu O, Su MY. Quantitative Analysis of Lesion Morphology and Texture Features for Diagnostic Prediction in Breast MRI. *Acad Radiol* 2008;15:1513–1525. [PubMed: 19000868]
6. Gweon HM, Cho N, Seo M, Chu AJ, Moon WK. Computer-aided evaluation as an adjunct to revised BI-RADS Atlas: improvement in positive predictive value at screening breast MRI. *Eur Radiol*. 2014;24(8):1800–1807. [PubMed: 24788228]

7. Newell D, Nie K, Chen JH, Hsu CC, Yu HJ, Nalcioglu O, Su MY. Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement. *Eur Radiol*. 2010;20(4):771–781. [PubMed: 19789878]
8. Gallego-Ortiz C, Martel AL. Improving the Accuracy of Computer-aided Diagnosis for Breast MR Imaging by Differentiating between Mass and Nonmass Lesions. *Radiology* 2016;278(3):679–688. [PubMed: 26383229]
9. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441–446. [PubMed: 22257792]
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563–577. [PubMed: 26579733]
11. Tsougos I, Vamvakas A, Kappas C, Fezoulidis I, Vassiou K. Application of Radiomics and Decision Support Systems for Breast MR Differential Diagnosis. *Comput Math Methods Med*. 2018;7417126. [PubMed: 30344618]
12. Chitalia RD, Kontos D. Role of texture analysis in breast MRI as a cancer biomarker: A review. *J Magn Reson Imaging* 2019;49(4):927–938. [PubMed: 30390383]
13. Li H, Zhu Y, Burnside ES, et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer* 2016;2:16012. [PubMed: 27853751]
14. Liang C, Cheng Z, Huang Y, et al. An MRI-based Radiomics Classifier for Preoperative Prediction of Ki-67 Status in Breast Cancer. *Acad Radiol* 2018;25:1111–1117. [PubMed: 29428211]
15. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500–510. [PubMed: 29777175]
16. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical Image Analysis using Convolutional Neural Networks: A Review. *J Med Syst*. 2018;42(11):226. [PubMed: 30298337]
17. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Lopez MAG. Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer methods and programs in biomedicine* 2016;127:248–257. [PubMed: 26826901]
18. Kooi T, Litjens G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*. 2017;35:303–312. [PubMed: 27497072]
19. Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Invest Radiol* 2017;52:434–440. [PubMed: 28212138]
20. Chougrad H, Zouaki H, Alheyane O. Deep Convolutional Neural Networks for Breast Cancer Screening. *Computer Methods and Programs in Biomedicine* 2018;157:19–30. [PubMed: 29477427]
21. Diniz JOB, Diniz PHB, Valente TLA, Silva AC, de Paiva AC, Gattass M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. *Computer Methods and Programs in Biomedicine* 2018; pii: S0169–2607(17)30445–5.
22. Al-Masni MA, Al-Antari MA, Park JM, et al. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Comput Methods Programs Biomed*. 2018;157:85–94. [PubMed: 29477437]
23. Antropova N, Abe H, Giger ML. Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. *Journal of Medical Imaging* 2018;5:014503. [PubMed: 29430478]
24. Haarbuerger C, Langenberg P, Truhn D, et al. Transfer learning for breast cancer malignancy classification based on dynamic contrast-enhanced MR images In: Maier A, Deserno T, Handels H, Maier-Hein K, Palm C, Tolxdorff T, eds. *Bild- verarbeitung für die Medizin* 2018. Berlin, Germany: Springer, 2018; 216–221.

25. Truhn D, Schrading S, Haarbuerger C, Schneider H, Merhof D, Kuhl C. Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-enhancing Lesions at Multiparametric Breast MRI. *Radiology*. 2019;290(2):290–297. [PubMed: 30422086]
26. Kim Y, Stolarska MA, Othmer HG. The role of the microenvironment in tumor growth and invasion. *Prog Biophys Mol Biol*. 2011;106(2):353–379. [PubMed: 21736894]
27. Wu JS, Sheng SR, Liang XH, Tang YL. The role of tumor microenvironment in collective tumor cell invasion. *Future Oncol*. 2017;13(11):991–1002. [PubMed: 28075171]
28. Lee A, DeLellis RA, Silverman ML, Heatley GJ, Wolfe HJ. Prognostic significance of peritumoral lymphatic and blood vessel invasion in node-negative carcinoma of the breast. *Journal of Clinical Oncology* 1990;8:1457–1465. [PubMed: 2202788]
29. Mohammed ZM, McMillan DC, Edwards J et al. The relationship between lymphovascular invasion and angiogenesis, hormone receptors, cell proliferation and survival in patients with primary operable invasive ductal breast cancer. *BMC clinical pathology* 2013;13:31. [PubMed: 24274633]
30. Freed M, Storey P, Lewin AA et al. Evaluation of breast lipid composition in patients with benign tissue and cancer by using multiple gradient-echo MR imaging. *Radiology* 2016;281:43–53. [PubMed: 27266558]
31. Shin HJ, Park JY, Shin KC et al. Characterization of tumor and adjacent peritumoral stroma in patients with breast cancer using high-resolution diffusion-weighted imaging: Correlation with pathologic biomarkers. *European journal of radiology* 2016;85:1004–1011. [PubMed: 27130063]
32. Cheon H, Kim HJ, Kim TH, et al. Invasive Breast Cancer: Prognostic Value of Peritumoral Edema Identified at Preoperative MR Imaging. *Radiology* 2018;287(1):68–75. [PubMed: 29315062]
33. Bezdek JC. *Objective Function Clustering in Pattern recognition with fuzzy objective function algorithms*: Springer, 1981; pp. 43–93.
34. Haralick RM and Shanmugam K. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 1973; no. 6, pp. 610–621
35. Ho TK. Random decision forests in Document analysis and recognition, proceedings of the third international conference IEEE 1995; vol. 1, pp. 278–282
36. Lang N, Zhang Y, Zhang E, et al. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. *Magn Reson Imaging*. 2019; pii: S0730–725X(18)30672–6. [Epub ahead of print]
37. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, June 27–30, 2016* Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2016; pp. 770–778
38. Shi L, Zhang Y, Nie K, et al. Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI. *Magn Reson Imaging*. 2019;61:33–40. [PubMed: 31059768]
39. Kingma D and Ba J. Adam: A method for stochastic optimization. arXiv preprint 2014;arXiv:1412.6980
40. Deng J, Dong W, Socher R, Li LJ, Li K, and Fei-Fei L. Imagenet: A large-scale hierarchical image database in *Computer Vision and Pattern Recognition, CVPR 2009. IEEE Conference, 2009*; pp. 248–255
41. Chai H, Brown RE. Field effect in cancer—an update. *Annals of Clinical & Laboratory Science* 2009;39:331–337 [PubMed: 19880759]
42. Heaphy CM, Griffith JK, Bisoffi M. Mammary field cancerization: molecular evidence and clinical importance. *Breast cancer research and treatment* 2009;118:229–239 [PubMed: 19685287]
43. Fan M, He T, Zhang P, Zhang J, Li L. Heterogeneity of Diffusion-Weighted Imaging in Tumours and the Surrounding Stroma for Prediction of Ki-67 Proliferation Status in Breast Cancer. *Sci Rep*. 2017;7(1):2875 [PubMed: 28588280]
44. Whitney HM, Taylor NS, Drukker K, et al. Additive Benefit of Radiomics Over Size Alone in the Distinction Between Benign Lesions and Luminal A Cancers on a Large Clinical Breast MRI Dataset. *Acad Radiol*. 2019;26(2):202–209 [PubMed: 29754995]

45. Antropova N, Huynh B, Li H, Giger ML. Breast lesion classification based on dynamic contrast-enhanced magnetic resonance images sequences with long short-term memory networks. *J Med Imaging (Bellingham)*. 2019 1;6(1):011002. doi: 10.1117/1.JMI.6.1.011002. [PubMed: 30840721]
46. Zhou J, Luo LY, Dou Q, Chen H, Chen C, Li GJ, Jiang ZF, Heng PA. Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. *J Magn Reson Imaging*. 2019 3 29. doi: 10.1002/jmri.26721. [Epub ahead of print]
47. Reig B, Heacock L, Geras KJ, Moy L. Machine learning in breast MRI. *J Magn Reson Imaging*. 2019 7 5. doi: 10.1002/jmri.26852. [Epub ahead of print]
48. Sheth D, Giger ML. Artificial intelligence in the interpretation of breast cancer on MRI. *J Magn Reson Imaging*. 2019 7 25. doi: 10.1002/jmri.26878. [Epub ahead of print]

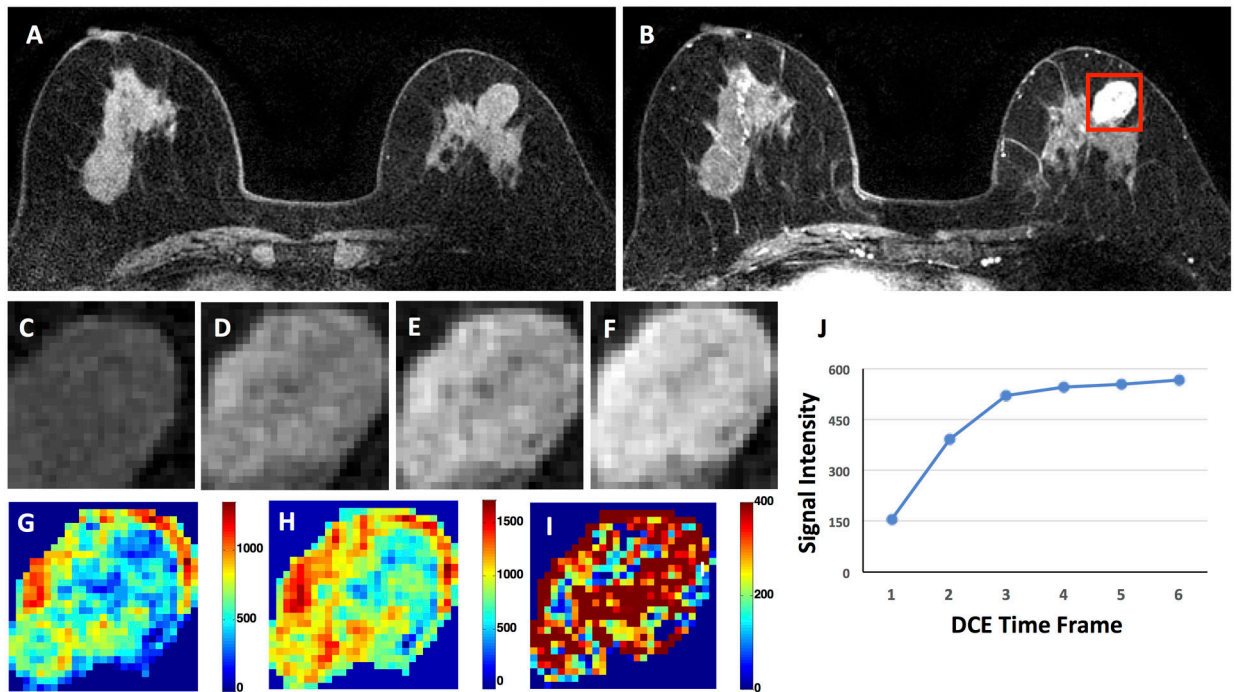


Figure 1:

A 66-year-old patient with a benign fibroadenoma showing smooth boundary. (A) F1 Pre-contrast image. (B) The F2 post-contrast image. The red square box is the smallest bounding box. (C-I): The zoom-in smallest bounding box containing the tumor. (C) The F1 pre-contrast image, (D) The F2 post-contrast image, (E) The F3 post-contrast image, (F) The last F6 post-contrast image, showing persistent enhancement with increased intensity over time. (G) The wash-in signal enhancement map F2-F1, (H) The F3-F1 signal enhancement map, (I) The wash-out F6-F3 map. (J) The DCE time course shows a persistent enhancement pattern from F1 to F6. The predicted malignancy probability is 0.69 for ROI-model (wrong), 0.20 for radiomics (correct), 0.23 for ROI+radiomics (correct), 0.36 for per-slice CNN (correct), 0.51 for per-lesion CNN (wrong based on threshold of 0.5). There are a total of 14 slices for this case, and only one slice has malignancy probability > 0.5 .

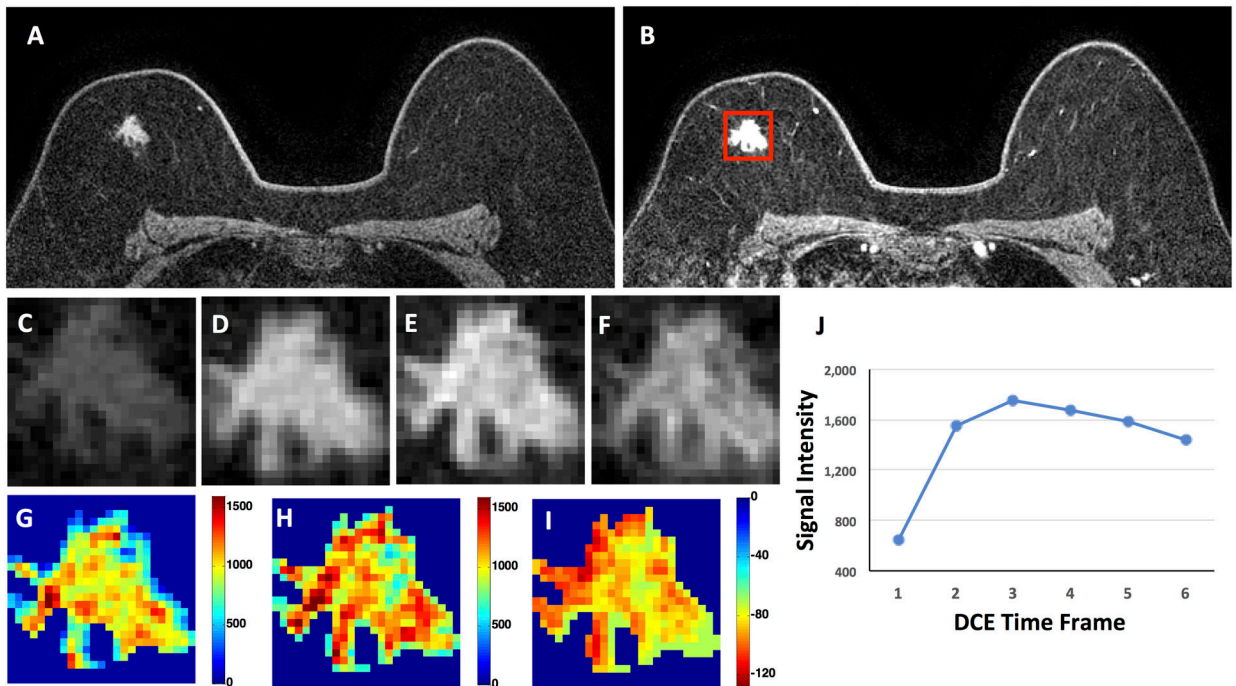


Figure 2:

A 68-year-old patient with a malignant invasive ductal cancer showing lobulated shape and spiculated margin. (A) F1 Pre-contrast image. (B) The F2 post-contrast image. The red square box is the smallest bounding box. (C-I): The zoom-in smallest bounding box containing the tumor. (C) The F1 pre-contrast image, (D) The F2 post-contrast image, (E) The F3 post-contrast image, (F) The last F6 post-contrast image, showing wash-out DCE pattern with decreased intensity after reaching maximum in F3. (G) The wash-in signal enhancement map F2-F1, (H) The maximum F3-F1 signal enhancement map, (I) The wash-out F6-F3 map. (J) The DCE time course shows a typical wash-out pattern, reaching maximum in F3, followed by decreased intensity from F4 to F6. The predicted malignancy probability is 0.83 for ROI-model, 0.97 for radiomics, 0.97 for ROI+radiomics, 0.97 for per-slice CNN, 0.99 for per-lesion CNN (all correct).

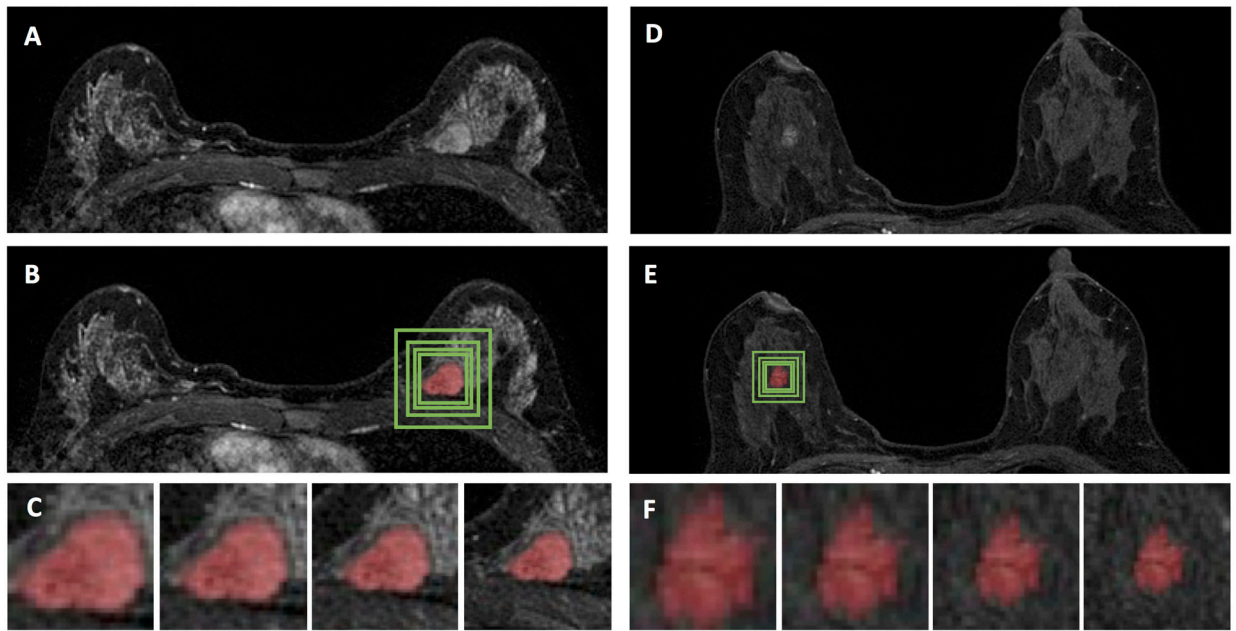


Figure 3:

Two benign cases. (A-C) A 41-year-old patient with a benign fibroadenoma showing smooth boundary. (A) The F3 post-contrast image. (B) The green box is the smallest square bounding box, and 1.2, 1.5, and 2 times expanded larger boxes. (C) The zoom-in image of the smallest, 1.2, 1.5, and 2 times boxes showing tumor with different amount of peri-tumor tissues. The predicted malignancy probability is 0.47 for ROI-model, 0.08 for radiomics, 0.10 for ROI+radiomics, 0.29 for per-slice CNN, 0.37 for per-lesion CNN (all correct). (D-F) A 54-year-old patient with a benign fibroadenoma showing low enhancement with indistinct boundary. The predicted malignancy probability is 0.28 for ROI-model, 0.02 for radiomics, 0.02 for ROI+radiomics, 0.29 for per-slice CNN, 0.29 for per-lesion CNN (all correct).

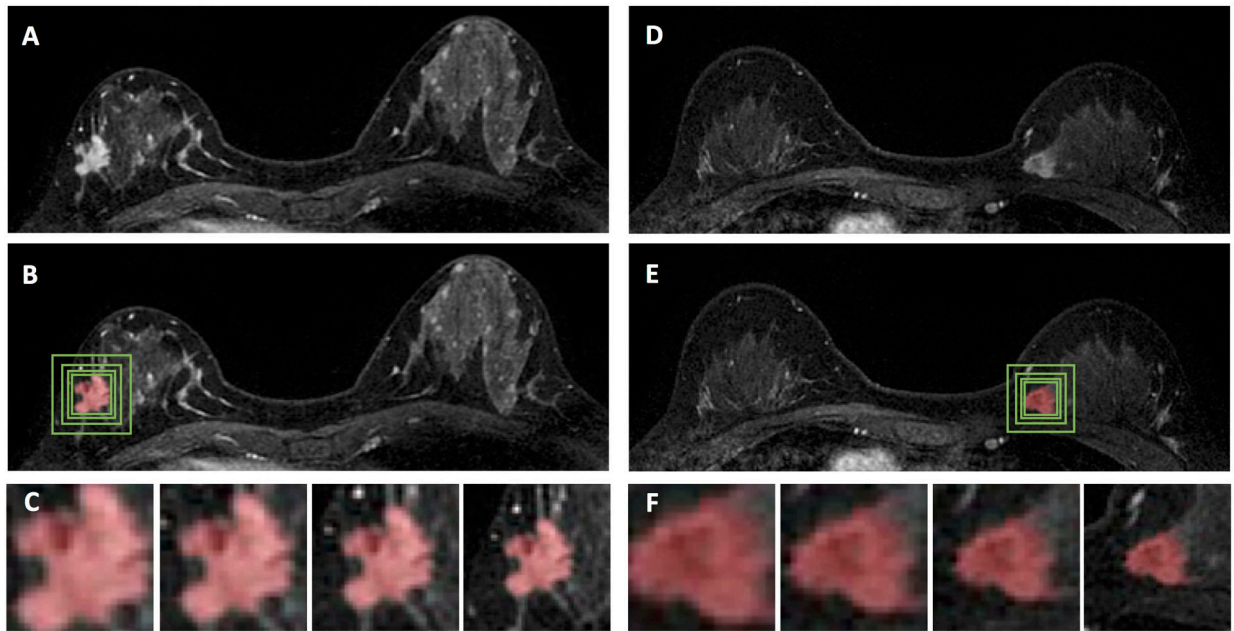


Figure 4:

Two malignant cases. (A-C) A 44-year-old patient with an invasive ductal cancer showing lobulated shape and spiculated margin. (A) The F3 post-contrast image. (B) The green box is the smallest square bounding box, and 1.2, 1.5, and 2 times expanded larger boxes. (C) The zoom-in image of the smallest, 1.2, 1.5, and 2 times boxes showing tumor with different amount of peri-tumor tissues. The predicted malignancy probability is 0.61 for ROI-model, 0.89 for radiomics, 0.90 for ROI+radiomics, 0.98 for per-slice CNN, 0.98 for per-lesion CNN (all correct). (D-F) A 41-year-old patient with an invasive ductal cancer with a clear medial boundary. The predicted malignancy probability is 0.41 for ROI-model, 0.29 for radiomics, 0.38 for ROI+radiomics (wrong prediction), 0.83 for per-slice CNN, 0.99 for per-lesion CNN (correct prediction).

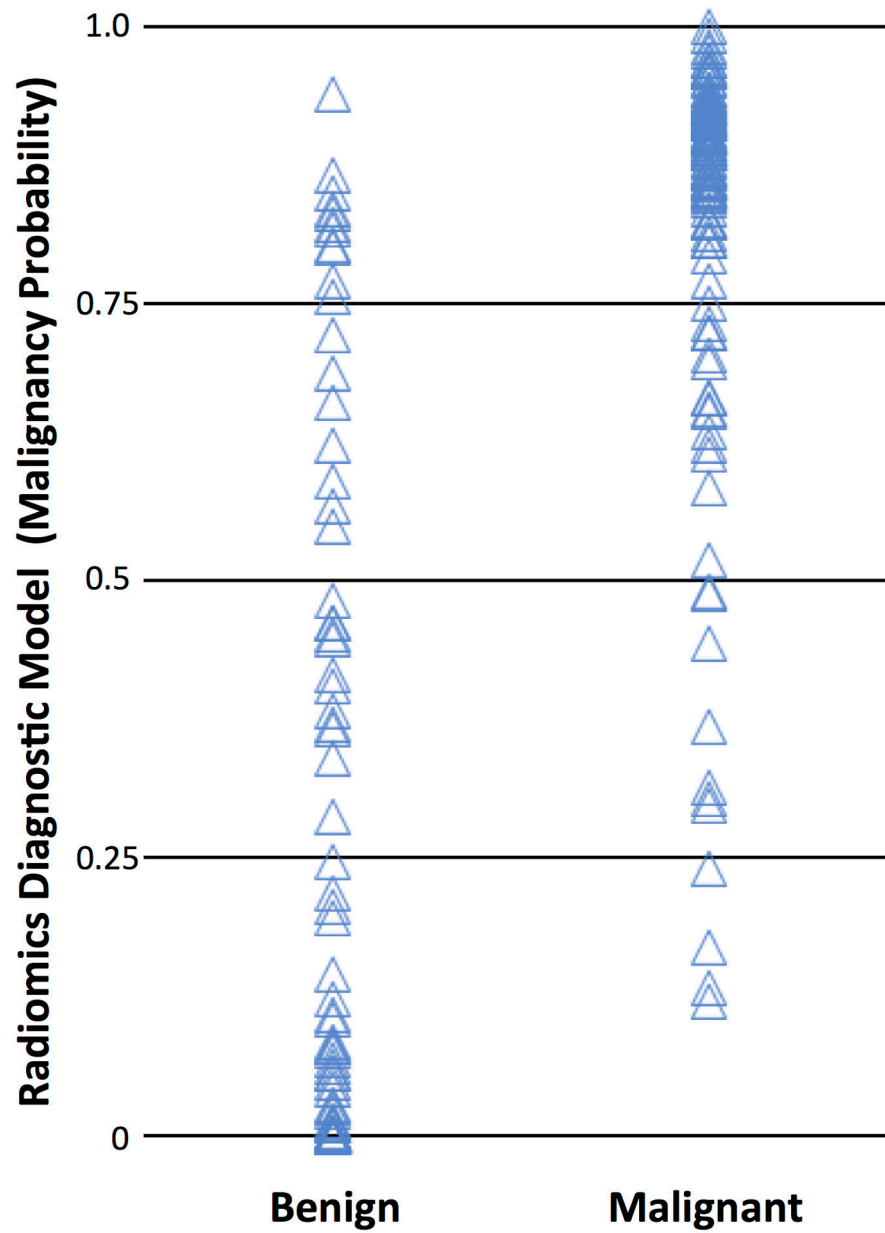


Figure 5: The plot of the malignancy probability calculated using the radiomics diagnostic model in the malignant and benign lesion groups. Based on the threshold of 0.5, the overall diagnostic accuracy is 84%. Of the total of 91 malignant and 62 benign cases, True Positive = 83 cases, True Negative = 45 cases, False Negative = 8 cases, False Positive = 17 cases.

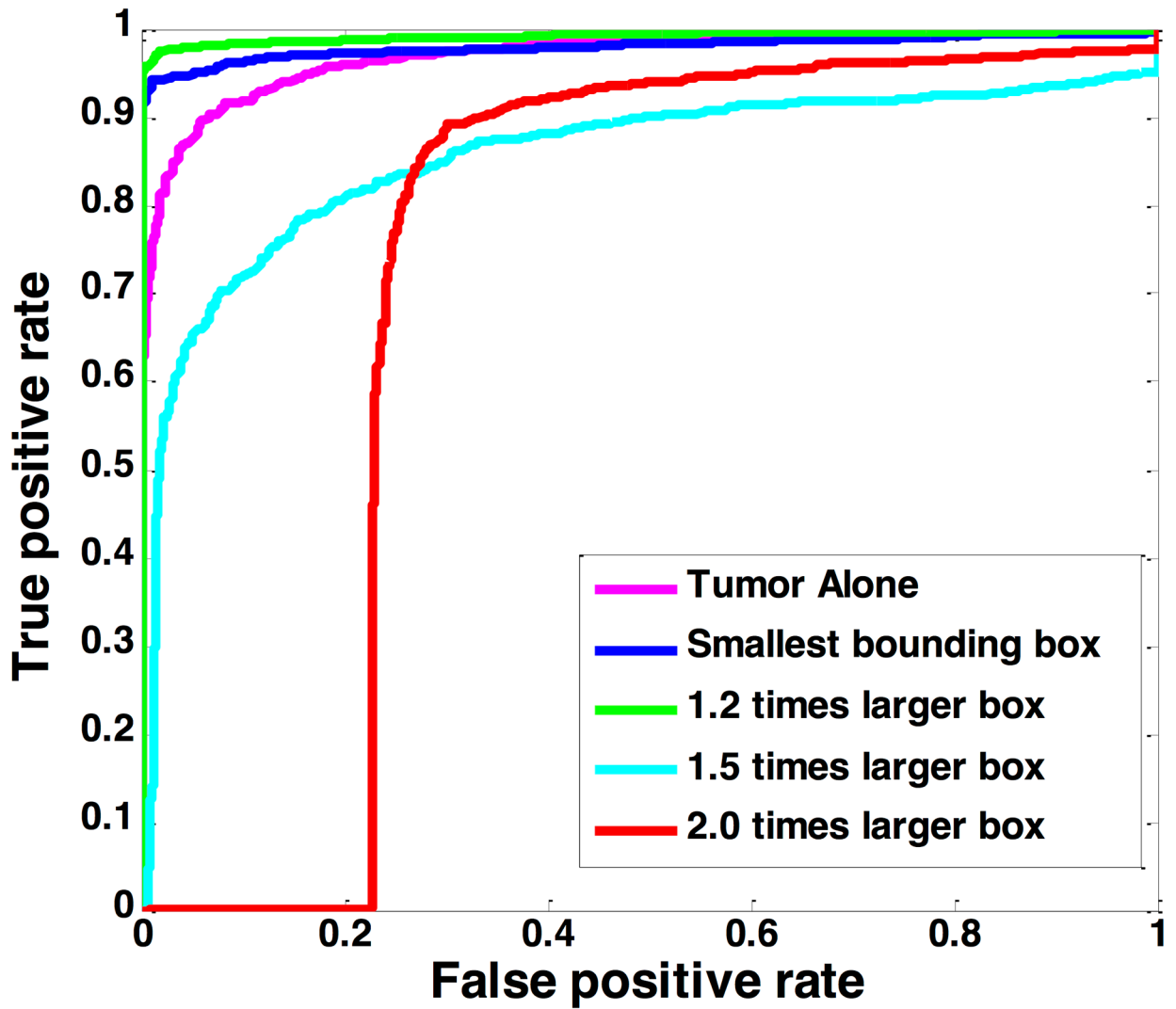


Figure 6: The ROC curves generated by using the predicted per-slice malignancy probability of the entire training dataset using ResNet50, with 5 different input methods: tumor alone, smallest bounding box, 1.2, 1.5, and 2.0 enlarged boxes.

Table 1.

The pathological subtypes in malignant and benign groups in training and testing datasets

Pathology Type	Training Dataset	Testing Dataset
Malignant	N=91	N=48
Invasive Ductal Cancer	75 (82%)	34 (70%)
Ductal Carcinoma In-Situ	11 (12%)	9 (20%)
Other Invasive Cancer	5 (6%)	5 (10%)
Benign	N=62	N=26
Adenosis	31 (50%)	13 (50%)
Fibroadenoma	15 (24%)	8 (32%)
Other Benign Lesions	16 (26%)	5 (18%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

The whole tumor ROI-based parameters in malignant and benign groups [mean \pm stdev]

	Training Dataset		Testing Dataset	
	Malignant (N=91)	Benign (N=62)	Malignant (N=48)	Benign (N=26)
Age	51 \pm 10	45 \pm 11	49 \pm 7	45 \pm 7
1-D size (cm) *	2.01 \pm 0.70	1.44 \pm 0.62	1.94 \pm 0.86	1.19 \pm 0.78
3D Volume (cm ³) *	3.74 \pm 3.09	1.09 \pm 1.46	4.16 \pm 3.25	1.13 \pm 1.60
Wash-in SE ratio *	1.61 \pm 0.80	1.15 \pm 0.65	1.43 \pm 0.75	1.22 \pm 0.83
Max SE ratio *	2.16 \pm 0.96	1.79 \pm 0.82	2.07 \pm 1.04	1.63 \pm 0.75
Wash-out slope *	-0.03 \pm 0.14	0.09 \pm 0.16	-0.02 \pm 0.12	0.05 \pm 0.09

* significantly different (p<0.05) between malignant and benign groups in both training dataset and testing dataset

Table 3.

The diagnostic sensitivity, specificity and the overall accuracy using models built by ROI-based volume and DCE parameters, radiomics, and ResNet50 deep learning, with a fixed threshold of malignancy probability=0.5

	Training Dataset (10-fold Cross-Validation)					Independent Dataset				
	Sensitivity %	Specificity %	Accuracy %	AUC	Sensitivity %	Specificity %	Accuracy %			
ROI Volume + DCE	77% (70/91)	74% (46/62)	76% (116/153)	0.82	71% (34/48)	62% (16/26)	67% (30/74)			
Radiomics	91% (83/91)	73% (45/62)	84% (128/153)	0.91	85% (41/48)	65% (17/26)	78% (58/74)			
ROI + Radiomics	91% (83/91)	77% (48/62)	86% (131/153)	0.91	83% (40/48)	65% (17/26)	77% (57/74)			
CNN, Per-Slice Basis										
ResNet (Tumor Alone)	95% (1285/1358)	74% (362/488)	89% (1647/1846)	0.97	84% (848/1022)	66% (190/289)	79% (1038/1311)			
ResNet (Smallest Box)	95% (1286/1358)	94% (460/488)	95% (1746/1846)	0.98	86% (879/1022)	79% (226/289)	84% (1105/1311)			
ResNet (1.2 Times Box)	99% (1338/1358)	86% (419/488)	95% (1757/1846)	0.99	78% (801/1022)	70% (202/289)	77% (1003/1311)			
ResNet (1.5 Times Box)	84% (1146/1358)	68% (334/488)	80% (1480/1846)	0.86	73% (741/1022)	66% (190/289)	71% (931/1311)			
ResNet (2.0 Times Box)	90% (1217/1358)	67% (326/488)	84% (1543/1846)	0.71	67% (687/1022)	59% (171/289)	65% (858/1311)			
CNN, Per-Lesion Basis										
ResNet (Tumor Alone)	100% (91/91)	61% (38/62)	84% (129/153)	N/A*	94% (45/48)	62% (16/26)	82% (61/74)			
ResNet (Smallest Box)	99% (90/91)	79% (49/62)	91% (139/153)	N/A	94% (45/48)	81% (21/26)	89% (66/74)			
ResNet (1.2 Times Box)	100% (91/91)	60% (37/62)	84% (128/153)	N/A	85% (41/48)	81% (21/26)	84% (62/74)			
ResNet (1.5 Times Box)	97% (88/91)	37% (23/62)	73% (112/153)	N/A	85% (41/48)	54% (14/26)	74% (55/74)			
ResNet (2.0 Times Box)	99% (90/91)	24% (15/62)	69% (105/153)	N/A	79% (38/48)	46% (12/26)	54% (40/74)			

* N/A: The highest malignancy probability in all slices of one lesion is used to make diagnosis for that lesion, with threshold of 0.5, thus not applicable for ROC analysis.

Table 4.

The per-lesion diagnostic results obtained using the model built by ResNet50 deep learning with the smallest bounding box, based on different threshold of malignancy probability varying from 0.5 to 0.7

Malignancy Probability	Training Dataset (91 Malignant, 62 Benign)			Testing Dataset (48 Malignant, 26 Benign)		
Threshold	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
0.50	99%	79%	91%	94%	81%	89%
0.55	98%	95%	97%	94%	81%	89%
0.60	98%	97%	97%	75%	92%	81%
0.65	98%	100%	99%	71%	92%	78%
0.70	95%	100%	97%	27%	100%	53%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript