UC Irvine
UC Irvine Previously Published Works

Title

Deep Learning-based Automatic Diagnosis of Breast Cancer on MRI Using Mask R-CNN for Detection Followed by ResNet50 for Classification

Permalink

https://escholarship.org/uc/item/9312g1kv

Journal

Academic Radiology, 30(Suppl 2)

ISSN

1076-6332

Authors

Zhang, Yang

Liu, Yan-Lin

Nie, Ke

et al.

Publication Date

2023-09-01

DOI

10.1016/j.acra.2022.12.038

Peer reviewed

Deep Learning-based Automatic Diagnosis of Breast Cancer on MRI Using Mask R-CNN for Detection Followed by ResNet50 for Classification

Yang Zhang, PhD,

Yan-Lin Liu, PhD,

Ke Nie, PhD,

Jiejie Zhou, MD, PhD,

Zhongwei Chen, MD,

Jeon-Hor Chen, MD,

Xiao Wang, PhD,

Bomi Kim, MD,

Ritesh Parajuli, MD,

Rita S. Mehta, MD,

Meihao Wang, MD, PhD,

Min-Ying Su, PhD

Department of Radiological Sciences, University of California, Irvine, California (Y.Z., Y.-L.L., J.-H.C., B.K., M.-Y.S.); Department of Radiation Oncology, Rutgers-Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, New Jersey (Y.Z., K.N., X.W.); Department of Radiology, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China (J.Z., Z.C., M.W.); Department of Radiology, E-Da Hospital and I-Shou University, Kaohsiung, Taiwan (J.-H.C.); Department of Breast Radiology, Ilsan Hospital, Goyang, South Korea (B.K.); Department of Medicine, University of California, Irvine, United States (R.P., R.S.M.); Department of Medical Imaging and Radiological Sciences, Kaohsiung Medical University, Kaohsiung, T (M.-Y.S.).

Abstract

Address correspondence to: M-Y. S. msu@uci.edu.

Rationale and Objectives: Diagnosis of breast cancer on MRI requires, first, the identification of suspicious lesions; second, the characterization to give a diagnostic impression. We implemented Mask Reginal-Convolutional Neural Network (R-CNN) to detect abnormal lesions, followed by ResNet50 to estimate the malignancy probability.

Materials and Methods: Two datasets were used. The first set had 176 cases, 103 cancer, and 73 benign. The second set had 84 cases, 53 cancer, and 31 benign. For detection, the pre-contrast image and the subtraction images of left and right breasts were used as inputs, so the symmetry could be considered. The detected suspicious area was characterized by ResNet50, using three DCE parametric maps as inputs. The results obtained using slice-based analyses were combined to give a lesion-based diagnosis.

Results: In the first dataset, 101 of 103 cancers were detected by Mask R-CNN as suspicious, and 99 of 101 were correctly classified by ResNet50 as cancer, with a sensitivity of 99/103 = 96%. 48 of 73 benign lesions and 131 normal areas were identified as suspicious. Following classification by ResNet50, only 16 benign and 16 normal areas remained as malignant. The second dataset was used for independent testing. The sensitivity was 43/53 = 81%. Of the total of 121 identified non-cancerous lesions, only 6 of 31 benign lesions and 22 normal tissues were classified as malignant.

Conclusion: ResNet50 could eliminate approximately 80% of false positives detected by Mask R-CNN. Combining Mask R-CNN and ResNet50 has the potential to develop a fully-automatic computer-aided diagnostic system for breast cancer on MRI.

INTRODUCTION

Breast cancer is the most commonly occurring cancer in women, accounting for 30% of the newly diagnosed female cancers in the United States in 2021 (1). The survival rate has greatly improved over the years through early detection and screening (2, 3). Breast magnetic resonance imaging (MRI) is a well-established imaging modality, which has been widely used for several clinical indications, including diagnosis, pre-operative staging, treatment response monitoring, and screening for women at high risk of developing cancer (4–6).

In the clinical workflow for radiologists' reading of breast MRI, the first task is to identify the suspicious abnormality, followed by characterization of the abnormal area(s) to make a diagnostic impression. Considering the large number of images acquired to cover the whole breast, it is time-consuming for radiologists to review the entire dataset. Computer-aided software has been developed as an assistance tool. The systems (e.g., Merge CADstream, DynaCAD) are mainly used to generate essential information to help radiologists interpret images. For example, the maximum intensity projection (MIP) of the subtraction images is the most useful for locating the enhanced areas in the entire volume, and thus, is often the first reviewed image. Then, the radiologist can review all images by paging through

slices to find the lesion and characterize it according to the Breast Imaging Reporting and Data System (BI-RADS) descriptors (7–9). The standard MRI protocol includes the dynamic-contrast-enhanced (DCE) MRI sequence to take multiple phases of post-contrast images, which can be used to generate the color-coded DCE wash-out maps and DCE time course. The information can be displayed together on the workstation for evaluation; however, all decisions must be made by the radiologist.

Benign and malignant lesions have distinct features that can be characterized using quantitative imaging parameters. In the last 3 decades, many computer-aided diagnosis (CAD) systems have been developed to make the differential diagnosis (7–13). Radiomics analysis can be applied to extract hundreds of features from the region of interest (ROI) of the lesion, and then sophisticated statistical methods, including machine learning algorithms, can be applied to select important features and build the best diagnostic model (14, 15). Although this approach can focus on the suspicious lesion, the ROI needs to be determined and segmented; thus, not easy to be implemented in clinical practice. Furthermore, experienced radiologists can achieve very high accuracy by visual assessment, and do not need to rely on quantitative features (16).

In recent years, deep learning has also been extensively applied to perform unsupervised analysis on breast MRI to detect and classify lesions, including differentiation of benign diseases and malignant cancers (16, 17), and prediction of the molecular subtypes (18). Convolutional Neural Network (CNN) is a popular architecture that can be applied to estimate the probability of malignancy for identified lesions (19, 20). It has been shown that the accuracy is dependent on the size of the input box; therefore, the location and the size of the suspicious lesion are needed to achieve a high accuracy (17). Deep learning has also been shown capable of searching the entire MRI dataset to detect abnormal lesions (21–23). Several CNN algorithms have been implemented, such as the Patch-based network (23–26) and Mask Regional-Convolutional Neural Network (R-CNN) (27, 28), to search the entire set of images or feature maps to detect and localize the lesion. However, these studies were mainly demonstrating feasibility, not deployed to test clinical datasets.

We have previously developed a fully-automatic detection model using Mask Reginal-Convolutional Neural Network (R-CNN) (28), but many enhanced areas were detected as false positives and they should be further characterized. We have also developed a diagnostic model using ResNet50 to predict the malignancy probability of the detected lesion (17). In the present study, the primary objective was to apply these two networks to test their detection and diagnostic performance in clinical datasets. In order to maintain high sensitivity, many false positive lesions will be detected, and the second objective was to categorize them to facilitate the future development of accurate methods to eliminate false positives.

MATERIALS AND METHODS

Patient Datasets

This study was conducted using two datasets. The first dataset (Dataset-1) had a total of 176 cases, including 103 malignant tumors (mean age of patients 55 ± 12) and 73 benign lesions

(mean age 42 ± 9). These cases were selected from consecutive patients receiving breast MRI for diagnosis from January 2017 to May 2018, before biopsy or any treatment. The majority of these cases were used in the training of our ResNet50 diagnostic model before (17). A second dataset (Dataset-2) was assembled using later cases collected from June 2018 to June 2019 as an independent testing dataset, with a total of 84 cases. The dataset included 53 malignant tumors (mean age of patients 56 ± 7) and 31 benign lesions (mean age 50 ± 9). All lesions, including benign diseases, had confirmed pathological diagnoses, and the major types are listed in Table 1. This study was approved by the Ethics Committee of the First Affiliated Hospital of Wenzhou Medical University, and the requirement of informed consent was waived.

MRI Protocol and Image Pre-processing

The MRI examination was performed using a 3T scanner (GE SIGNA HDx) with a dedicated 8-channel bilateral breast coil. The volume imaging for breast assessment (VIBRANT) sequence was used for the DCE acquisition, in the axial view covering both breasts. The imaging parameters were: repetition time (TR) = 5 msec; echo time (TE) = 2 msec; flip angle (FA) = 10°; slice thickness = 1.2 mm; field of view (FOV) = 34 $\times$ 34 cm^2; matrix size = 416 $\times$ 416. A total of 116 images were acquired to cover the whole breast. The DCE series consisted of six frames: one pre-contrast (F1) and five post-contrast (F2-F6). The acquisition time for each frame was 1 minute 32 seconds. The contrast agent, 0.1 mmol/kg Gadopentetate Dimeglumine (Magnevist; Bayer Schering Pharma), was intravenously injected after the pre-contrast images were acquired, at a rate of 2 mL/s, followed by a 20-mL saline flush at the same rate.

The analysis flowchart is shown in Figure 1. The detection and characterization networks can be performed separately, and don't need to be connected. First, the Mask R-CNN is used for detection, which will find suspicious lesions in bounding boxes. Then each suspicious bounding box is used as the input into the ResNet50 model to calculate the malignancy probability. For detection, we used breast symmetry to serve as a reference, so the image was split into half from the mid-line to separate the left breast and the right breast. For classification, the images acquired in the 6 DCE frames were used to generate three heuristic parametric maps:

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

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

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

These three maps were used as input into ResNet50. The malignant lesions usually reached the maximum signal intensity around F3 (at 2–3 minutes after contrast injection). For benign lesions, although most of them would not show a maximum at F3, the maps generated using

the same rule could yield distinctly different features to be distinguished from malignant lesions.

Detection of Suspicious Areas Using Mask R-CNN

Mask R-CNN Architecture—The deep learning detection algorithm was based on a custom Mask R-CNN architecture, shown in Figure 2. The network design was described before, and the detailed methods can be found there (28). Various predefined shapes and distributions of bounding boxes were generated in the entire image to identify potential abnormalities, and then they were ranked based on the likelihood. Those with the highest probabilities were extracted to generate region proposals to locate specific regions, pruned using non-maximum suppression, and put into a classifier to determine whether these regions belonged to lesion or non-lesion. For the detected lesion, a bounding box was generated, and further, a segmentation network could be added. ResNet101 was used as the feature pyramid network to work as the backbone, which contains one 3×3 convolutional layer, one max pooling layer, and 48 residual blocks. Each block contains one 1×1 convolutional layer, one 3×3 convolutional layer, and one 1×1 convolutional layer. Since the network was based on the ResNet, which was pre-trained using natural images with the RGB color, the allowed number of input channels was 3. The inputs from the feature pyramid network bottom-up pathway were added to the feature maps using a projection operation to match matrix dimensions, as shown in Figure 2.

Previous Training and Validation Methods—The training and testing of the Mask R-CNN model were reported before (28). The analysis was performed using the second post-contrast images (F3, acquired at 2–3 minutes after injection) when the maximum signal intensity was reached in malignant lesions. It was trained using 241 malignant cases presenting as mass lesions, acquired using the non-fat-saturated DCE sequence. The trained model was applied to 98 malignant cases acquired using the fat-saturated DCE sequence for testing. The results showed that 240 of 241 (99.5%) lesions in the training and 98 of 98 (100%) lesions in the testing datasets were correctly identified.

Using the Trained Mask R-CNN to Identify Suspicious Lesions—The images in the present study were acquired using a fat-saturated DCE sequence, which was the same as the images used in the testing dataset before and proven to work. The images were split along the mid-line to separate the left and right breasts. Three images (F1 pre-contrast and F3 subtraction images of the left and right breasts) were used as the input, and the network would identify suspicious lesions, each in a bounding box, as the output. From the previous training, it was known that the Mask R-CNN model could detect cancers, but it also detected many false positive areas, which should be further evaluated to rule them out. We used a previously developed ResNet50 model for classification, as described below.

Classification Using ResNet50

ResNet50 Architecture—The output bounding box of the abnormal area detected by Mask R-CNN was used to crop the heuristic DCE parametric maps and put them into the ResNet50 for classification. The ResNet50 Model used here was trained and tested in our previous study (17), as shown in Figure 3. The architecture contains 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 a sigmoid function to give the probability for 2-way prediction of malignant versus benign.

Previous Training and Validation Methods—Each tumor slice of a lesion was used as an independent input. The bounding box was resized to 75×75 pixels as input into the networks. This matrix was decided based on the range of tumor sizes in the analyzed datasets, from 10 to 180 pixels, so a median value of 75 pixels was chosen to standardize the input box. Random affine transformation was applied to augment the dataset to 20 times. The training was implemented using the Adam optimizer, which is a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments. It is computationally efficient and suitable for complicated models with a large number of trainable parameters. The learning rate was set at 0.001. The loss function was cross-entropy, which is the most widely used loss function in classification applications. It measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label. Parameters were initialized using ImageNet. The L2 regularization was performed to prevent overfitting. The ResNet50 model was trained using histologically confirmed 91 malignant and 62 benign mass lesions, and tested using 48 malignant and 26 benign lesions. The lesion-based diagnostic accuracy was 91% for training cases, and 89% for the testing cases.

Using the Trained ResNet50 to Perform Classification—Each identified lesion from Mask R-CNN was characterized by ResNet50 to estimate the malignancy probability. The analysis was performed using each 2D slice as an independent input, which meant that each slice had its own diagnostic probability. For lesion-based diagnosis, the probabilities of all slices of one lesion were considered. If the probability was 0.5 in more than 3 slices, the lesion was considered malignant. Lesions smaller than 4 mm were unlikely to be malignant. Therefore, with 1.2 mm slice thickness, if only 3 or fewer consecutive images of a lesion showed malignant predictions, it was determined as negative.

Identification of the Source of False Positives

For each lesion with a histologically confirmed diagnosis, the location and size were known and documented. For all Mask R-CNN identified lesions that were determined to be malignant by ResNet50, we evaluated their location and size, and then used the documented histology record to determine whether they were true positive cancers or false positive findings from confirmed benign lesions. Normal vessels or tissues showing strong contrast enhancements could be identified by Mask R-CNN as false positives. We further reviewed all wrongly predicted areas/lesions to determine whether they were coming from vessels or normal parenchymal enhancements. For determining the false ROI from vessels, the subtraction images of the second frame (F2, first post-contrast images) were used to generate the maximum intensity projection (MIP) to visualize the vessels. If the detected ROI fell on the vascular tree, it was identified as coming from vessels. For the ROI's related to parenchymal enhancements, the area was usually irregular, larger, without clear boundary,

and displayed persistent enhancements in the DCE series. Therefore, they could be better determined from the later DCE frame post-contrast images (e.g., F5 or F6).

Statistical Analysis

For each Mask R-CNN identified lesion, the malignancy probabilities of all slices contained in this lesion were estimated using the ResNet50 model. Then, if more than three slices were showing a probability 0.5, the lesion was determined as malignant. The true positive (TP) and false positive (FP) cases in Dataset-1 and Dataset-2 were calculated. Also, since the number of malignant cases was known, the false negative (FN) cases were evaluated. The sensitivity was calculated as TP/(TP+FN). For histologically confirmed benign lesions, the specificity could be calculated by TN/(TN+FP from benign lesions). The positive predictive value (PPV, or Precision) was calculated as TP/(TP+Total FP). The False Detection Rate (FDR) was calculated as FP/(FP+TP).

RESULTS

Detection and Classification in Dataset-1

All cases in Dataset-1 were mass lesions. Mask R-CNN identified 101 of the 103 malignant lesions and missed 2 small lesions. In 73 benign lesions, 48 were identified by Mask R-CNN. Also, 131 normal areas were detected as suspicious. All these detected areas were put into the ResNet50 to evaluate the malignancy probability. In the malignancy group, 99/101 detected lesions were correctly diagnosed as malignant. In the benign group, 16/48 detected lesions were diagnosed as malignant. In the detected 131 normal areas, only 16 were mis-diagnosed as malignant. Among the total of 179 false positives, only 32 remained as malignant after classification, so (179-32)/179 = 82% were eliminated. Combining the detection and diagnosis results together, 99 of the 103 malignant cancers were diagnosed as true positive (TP), and 4 were missed as false negative (FN, including 2 not detected, and 2 detected but diagnosed as benign). The diagnostic sensitivity was 96.1%. There were a total of 32 false positive (FP) cases, 16 from confirmed benign cases, and 16 from normal tissues. For the 16 FP from normal tissues, nine were from the parenchymal enhancements, and seven were from vessels. The positive predictive value was 99/(99+32) = 75.6%. The diagnostic results are summarized in Table 2.

Detection and Classification in the Independent Dataset-2

Dataset-2 was a completely independent dataset not used in any training before, and it also contained non-mass lesions. In 53 malignant lesions, Mask R-CNN correctly detected 49 and missed 4 small lesions. In 31 benign lesions, 27 were identified by Mask R-CNN. Also, 94 normal areas were detected as suspicious. All these detected areas were evaluated by ResNet50. In the malignancy group, 43/49 detected lesions were correctly diagnosed as malignant. Six of 27 detected benign lesions were diagnosed as malignant. In the detected 94 normal areas, 22 were mis-diagnosed as malignant. Therefore, among the total 121 false positives, only 28 remained as malignant after classification, so (121-28)/121 = 77% were eliminated. Combining the detection and diagnosis results together, 43/53 malignant cancers were diagnosed as true positive (TP), and 10 were missed (FN, including four not detected, and six detected but diagnosed as benign). The diagnostic sensitivity was 81.1%. There

were a total of 28 false positive (FP) cases, including 6 from confirmed benign cases and 22 from normal tissues. For the 22 FP from normal tissue, 14 were from the parenchymal enhancements, and eight were from vessels. The positive predictive value was 43/(43+28) = 60.6%.

Detection and Classification of Non-Mass Lesions in Dataset-2

While Dataset-1 contained all mass lesions, in Dataset-2, 26 of 84 cases were non-mass enhancement (NME) lesions, and 20 of 26 were malignant. Of them, 14 lesions were determined as suspicious. The sensitivity of detecting the malignant NME was 70% (14/20), lower than the sensitivity of 81% in the entire group. However, 5 out of 6 confirmed benign NME were detected as false positives, suggesting that our method has the capability of detecting the suspicious benign lesions that warrant biopsy (e.g., inflammation, and regional adenosis that appears as ductal carcinoma in situ - DCIS).

True Positive, True Negative, and False Positive Examples

Several case examples showing TP, TN, and FP diagnoses are illustrated. Figure 4 shows a true positive malignant diagnosis with a high probability of 0.92 for an invasive ductal cancer. Figure 5 shows an interesting case, in which two abnormal ROI's are identified by the Mask R-CNN as suspicious, and both are correctly determined as negative with a probability <0.5. One lesion is a histologically confirmed adenosis with a probability of 0.44; and the other is from parenchymal enhancement with a low probability of 0.12. Figure 6 shows several false positive cases from histologically confirmed benign lesions. They present as strongly enhanced mass lesions with a smooth boundary, a characteristic feature of benign lesions, especially for fibroadenomas. Figure 7 shows several false positive cases from vessels, which can be confirmed from the corresponding MIP that demonstrates the large vascular trees. Figure 8 shows several false positive cases from asymmetric parenchymal enhancements. These false positive ROI's from vessels and normal parenchymal enhancements can be easily ruled out by radiologists or using other computer algorithms.

DISCUSSION

In this study, we combined two deep learning networks, first to identify suspicious areas using Mask R-CNN, and then to classify them as benign or malignant using ResNet50. The Dataset-1 was not used in prior training for the Mask R-CNN detection, so it could be used to test the performance of the detection model. The combined networks achieved a high sensitivity of 96% and a high positive predictive value of 76%. The second dataset was totally new, thus it could be used for independent testing. There were more non-mass enhancing lesions in Dataset-2, which was much more challenging for detection and diagnosis than mass lesions (28). The sensitivity of 81% and the positive predictive value of 61% were lower than that in Dataset-1, but these can be adjusted by varying the threshold. Therefore, although Mask R-CNN identified many suspicious areas not related to cancer, when ResNet50 was utilized to characterize them, approximately 80% of wrongly identified areas could be dismissed to reduce the false positive while maintaining a high true positive detection.

Breast cancer has heterogeneous appearances on MRI, ranging from obvious masses with strong enhancements and spiculated margins to subtle asymmetry with mild enhancements (e.g., low grade ductal carcinoma in-situ), leading to difficulties for making accurate diagnosis and consistent interpretation. Many studies have investigated the value of machine learning for differentiation of benign and malignant lesions, by using the radiomics analysis or deep learning to characterize the identified abnormal lesions. However, the detection of suspicious abnormal areas was a much more challenging task, especially in MRI where many images were acquired to cover the entire breast with very thin slices (e.g., 1.2 mm in our DCE protocol). One main goal of the deep-learning in medical imaging is to build efficient tools for labeling the abnormal areas, to assist physicians and improve their performance and efficiency.

Deep learning has been shown capable of searching and detecting abnormalities on pathological whole-slide images (29). For breast lesions, most detection studies were developed for 2-dimensional mammography (25), and the algorithms can be applied to digital breast tomosynthesis (DBT) (26). In addition to the patch-based method, another feasible method is the weakly supervised learning. Kim *et al.* (30) analyzed the four-view digital mammograms, similar to the reading of radiologists in clinics. The feature maps before the global pooling layers were extracted to generate the probability maps, which could identify the detected lesion location and further to indicate the level of suspicion by using heat maps with different colors. Lu *et al.* (31) designed a CNN model based on ResNet18 and spatial attention to detect breast cancers on ultrasound images. Gao *et al.* (32) used another ResNet based architecture, shallow-deep CNN, to detect lesions on mammography and contrast enhanced mammography. Ribli *et al.* (27) implemented the faster R-CNN algorithm using VGG16 as a backbone network and reported that the system could detect and classify malignant or benign lesions on a mammogram without any human intervention. The streamlined procedure has been implemented in commercial products for mammography and DBT.

For lesion detection on breast MRI, because many images were acquired with different pulse sequences, it was much more challenging compared to the detection on other imaging modalities. Deep learning has provided a feasible method. Dalmı *et al.* (23) developed a computer aided detection system from the early-phase DCE scans. The system used 3D morphological information in the candidate locations and the enhancement differences of the two breasts, also considering symmetry. Free response ROC curves were used for the evaluation of the cancer detection rates with respect to the allowed false positives. Herent *et al.* (33) implemented a deep learning framework using ResNet50 for detection as well as diagnosis of four tissue categories: mammary gland, benign lesions, invasive ductal carcinoma, other malignant lesions. Respective ROC for each tissue type was constructed, which showed a weighted mean AUC of 0.816. However, this patch-based method would lead to high false positive rates, which were not further analyzed. Zhou *et al.* (34) implemented a weakly supervised learning method using Dense-Net to differentiate breast cancer. Gradient-weighted class activation mapping (Grad-CAM) was used to localize the suspicious areas. The detection results reached 83.7% accuracy, 90.8% sensitivity, and 69.3% specificity. Jing *et al.* (35) designed a ResNet34 algorithm to detect and classify breast cancers in ultrafast breast MRI datasets. The trained model achieved an AUC of

0.81, and a high sensitivity and negative predictive value could be achieved by adjusting the threshold level. Ayatollahi *et al.* (36) applied RetinaNet to detect breast lesions using all phases of DCE MRI, which could extract 4D (3D + time) information. When allowing four false positives per normal breast, a high detection rate of 0.95 for cancer and 0.81 for benign lesions could be achieved. These studies were mainly focused on the detection of known lesions, without further analysis of false positives in normal tissues. Moreover, different datasets, for example, diagnostic versus screening studies, will have very different sensitivity and specificity, and difficult to make fair comparisons.

For most object detection algorithms, a high sensitivity is associated with high false positives. As shown in our previous study (28), the number of FPs can be 1.5 to 2 times higher than the number of TPs in an all-malignant dataset. When benign cases were added, false positive rates would be expected to be higher. After the lesion was identified, it could be further segmented and characterized to make a diagnosis as benign or malignant. Our results showed that ResNet50 could be applied to eliminate 240 false positives from a total 300, approximately 80% (from 179 to 32 in Dataset-1, and from 121 to 28 in Dataset-2).

This study has several limitations. First, although two datasets were used, they were from the same hospital and the images were acquired using the same MR scanner during different time periods. Second, our ResNet50 diagnostic model was developed for mass lesions, which might explain the lower accuracy for Dataset-2. The model needs to be further improved for diagnosis of non-mass lesions (37). Third, we have shown that many false positives are from the vessels, which can be identified based on the morphological operations, for example, using the algorithm to detect linear structures on vascular MIP (38). On the other hand, the normal parenchymal enhancements are difficult to be evaluated by computer algorithms and will probably need radiologists' evaluation to eliminate them. Lastly, the study was performed using two previously developed models, not aiming to develop more accurate algorithms. We have tried to implement YOLO, YOLO3, and Faster R-CNN for lesion detection, but found them inferior to Mask R-CNN. More sophisticated deep learning algorithms may be implemented to improve performance.

CONCLUSION

We implemented a fully-automatic deep learning method, by first applying the Mask R-CNN to search the entire MRI dataset and identify suspicious lesions, followed by ResNet50 to characterize the lesions, and predict the diagnosis as malignant or benign. The results in two datasets suggest that high sensitivity can be achieved, and that many false positives detected by Mask R-CNN can be eliminated by ResNet50. For the false positives coming from vessels and the asymmetric parenchymal enhancements, these can be quickly discarded by radiologists, or by using other computer algorithms according to their unique morphological features. The results suggest that the proposed Mask R-CNN and ResNet50 have the potential to provide a fully-automatic computer-aided diagnosis system for breast MRI.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021; 71(1):7–33. [PubMed: 33433946]

2. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 Guideline Update From the American Cancer Society. JAMA 2015; 314(15):1599–1614. [PubMed: 26501536]

3. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin 2007; 57(2):75–89. [PubMed: 17392385]

4. Raikhlin A, Curpen B, Warner E, et al. Breast MRI as an adjunct to mammography for breast cancer screening in high-risk patients: retrospective review. AJR Am J Roentgenol 2015; 204(4):889–897. [PubMed: 25794083]

5. Marino MA, Helbich T, Baltzer P, Pinker-Domenig K. Multiparametric MRI of the breast: A review. J Magn Reson Imaging 2018; 47(2):301–315. [PubMed: 28639300]

6. Mann RM, Kuhl CK, Moy L. Contrast-enhanced MRI for breast cancer screening. J Magn Reson Imaging 2019; 50(2):377–390. [PubMed: 30659696]

7. 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]

8. Nie K, Chen JH, Yu HJ, et al. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. Acad Radiol 2008; 15(12):1513–1525. [PubMed: 19000868]

9. 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]

10. Newell D, Nie K, Chen JH, et al. 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]

11. Cho N, Kim SM, Park JS, et al. Contralateral lesions detected by preoperative MRI in patients with recently diagnosed breast cancer: application of MR CAD in differentiation of benign and malignant lesions. Eur J Radiol 2012; 81(7):1520–1526. [PubMed: 21498016]

12. Eun NL, Son EJ, Gweon HM, Youk JH, Kim JA. The value of breast MRI for BI-RADS category 4B mammographic microcalcification: based on the 5th edition of BI-RADS. Clin Radiol 2018; 73(8):750–755. [PubMed: 29853301]

13. 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]

14. 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(4):441–446. [PubMed: 22257792]

15. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278(2):563–577. [PubMed: 26579733]

16. Truhn D, Schrading S, Haarburger C, et al. Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. Radiology 2019; 290(2):290–297. [PubMed: 30422086]

17. Zhou J, Zhang Y, Chang KT, et al. Diagnosis of benign and malignant breast lesions on DCE-MRI by using radiomics and deep learning with consideration of peritumor tissue. J Magn Reson Imaging 2020; 51 (3):798–809. [PubMed: 31675151]

18. Zhang Y, Chen JH, Lin Y, et al. Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. Eur Radiol 2021; 31(4):2559–2567. [PubMed: 33001309]

19. Lee JG, Jun S, Cho YW, et al. Deep learning in medical imaging: general overview. Korean J Radiol 2017; 18(4):570–584. [PubMed: 28670152]

20. Al-Masni MA, Al-Antari MA, Park JM, et al. Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network. Annu Int Conf IEEE Eng Med Biol Soc 2017; 2017:1230–1233. [PubMed: 29060098]

21. Codari M, Schiaffino S, Sardanelli F, Trimboli RM. Artificial intelligence for breast MRI in 2008–2018: a systematic mapping review. AJR Am J Roentgenol 2019; 212(2):280–292. [PubMed: 30601029]

22. Sheth D, Giger ML. Artificial intelligence in the interpretation of breast cancer on MRI. J Magn Reson Imaging 2020; 51(5):1310–1324. [PubMed: 31343790]

23. Dalmı MU, Vreemann S, Kooi T, et al. Fully automated detection of breast cancer in screening MRI using convolutional neural networks. J Med Imaging (Bellingham) 2018; 5(1):014502. [PubMed: 29340287]

24. Yap MH, Pons G, Marti J, et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform 2018; 22(4):1218–1226. [PubMed: 28796627]

25. 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]

26. Samala RK, Chan HP, Hadjiiski L, et al. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Med Phys 2016; 43(12):6654. [PubMed: 27908154]

27. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep 2018; 8(1):4165. [PubMed: 29545529]

28. Zhang Y, Chan S, Park VY, et al. Automatic detection and segmentation of breast cancer on MRI Using Mask R-CNN trained on non-fat-sat images and tested on fat-sat images. Acad Radiol 2022; 29(Suppl 1): S135–S144. [PubMed: 33317911]

29. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017; 318(22):2199–2210. [PubMed: 29234806]

30. Kim EK, Kim HE, Han K, et al. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. Sci Rep 2018; 8(1):2762. [PubMed: 29426948]

31. Lu SY, Wang SH, Zhang YD. SAFNet: A deep spatial attention network with classifier fusion for breast cancer detection. Comput Biol Med 2022; 148:105812. [PubMed: 35834967]

32. Gao F, Wu T, Li J, et al. SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Comput Med Imaging Graph 2018; 70:53–62. [PubMed: 30292910]

33. Herent P, Schmauch B, Jehanno P, et al. Detection and characterization of MRI breast lesions using deep learning. Diagn Interv Imaging 2019; 100(4):219–225. [PubMed: 30926444]

34. Zhou J, Luo LY, Dou Q, et al. Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J Magn Reson Imaging 2019; 50(4):1144–1151. [PubMed: 30924997]

35. Jing X, Wielema M, Cornelissen LJ, et al. Using deep learning to safely exclude lesions with only ultrafast breast MRI to shorten acquisition and reading time. Eur Radiol 2022. doi:10.1007/s00330-022-08863-8.

36. Ayatollahi F, Shokouhi SB, Mann RM, Teuwen J. Automatic breast lesion detection in ultrafast DCE-MRI using deep learning. Med Phys 2021; 48 (10):5897–5907. [PubMed: 34370886]

37. Zhou J, Liu YL, Zhang Y, et al. BI-RADS reading of non-mass lesions on dce-mri and differential diagnosis performed by radiomics and deep learning. Front Oncol 2021; 11:728224. [PubMed: 34790569]

38. Lin M, Chen JH, Nie K, et al. Algorithm-based method for detection of blood vessels in breast MRI for development of computer-aided diagnosis. J Magn Reson Imaging 2009; 30(4):817–824. [PubMed: 19787727]
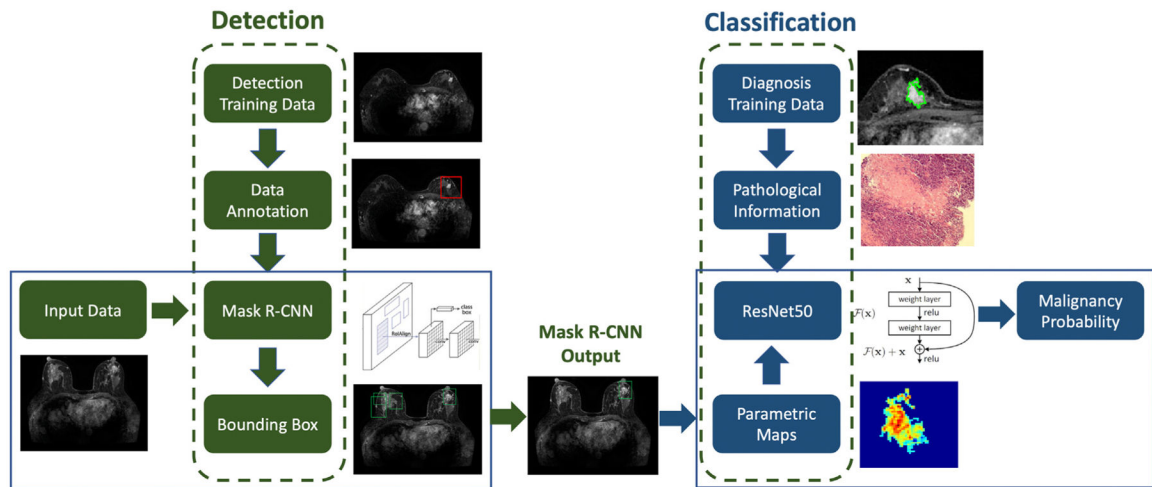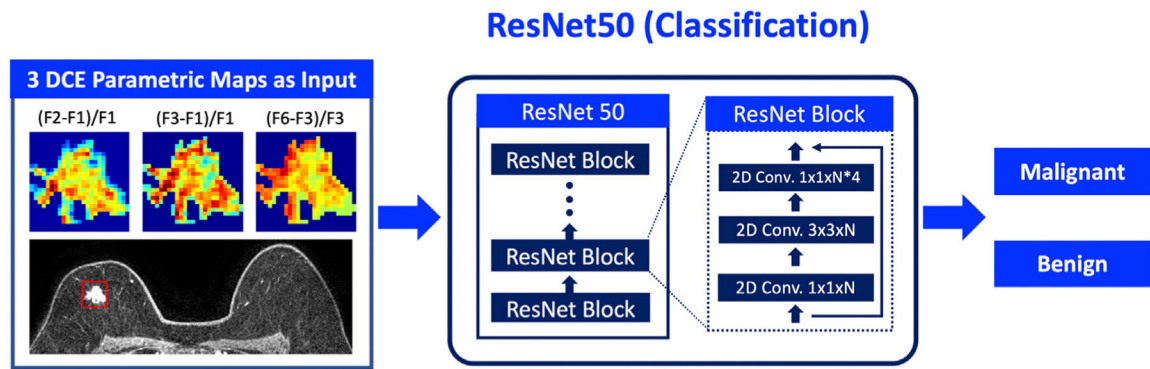
Figure 1.
Analysis flowchart using Mask R-CNN for detection of suspicious abnormal areas, followed by ResNet50 for classification of each detected abnormal area into malignant or benign as a diagnostic prediction. These two-step procedures can be performed separately. The output of the suspicious lesion bounding box by Mask R-CNN is used as the input into ResNet50. The Mask R-CNN and ResNet50 models have been trained and tested before, so the analysis procedures for the present study are in the marked bottom half.
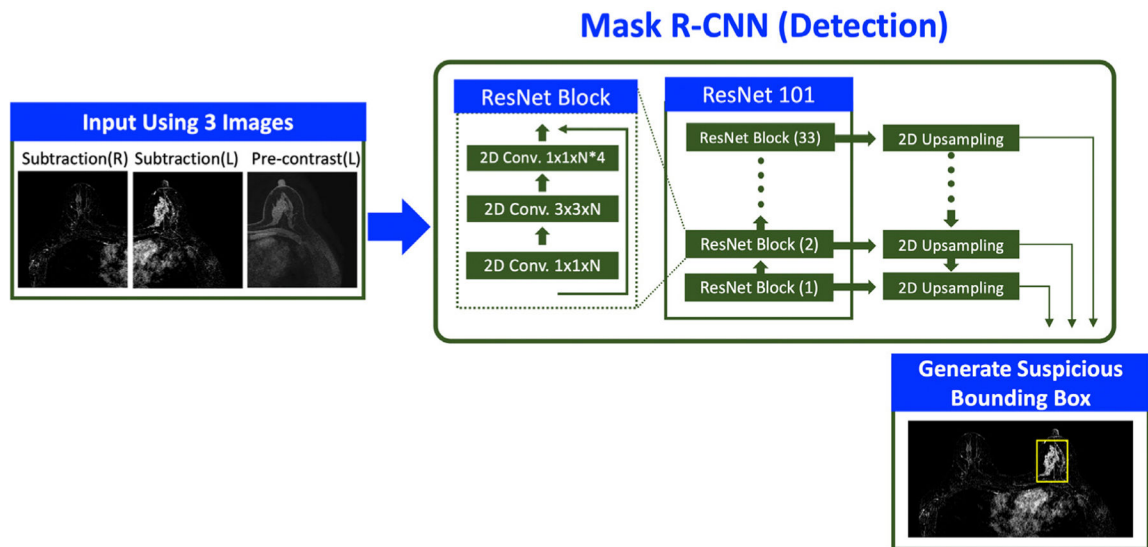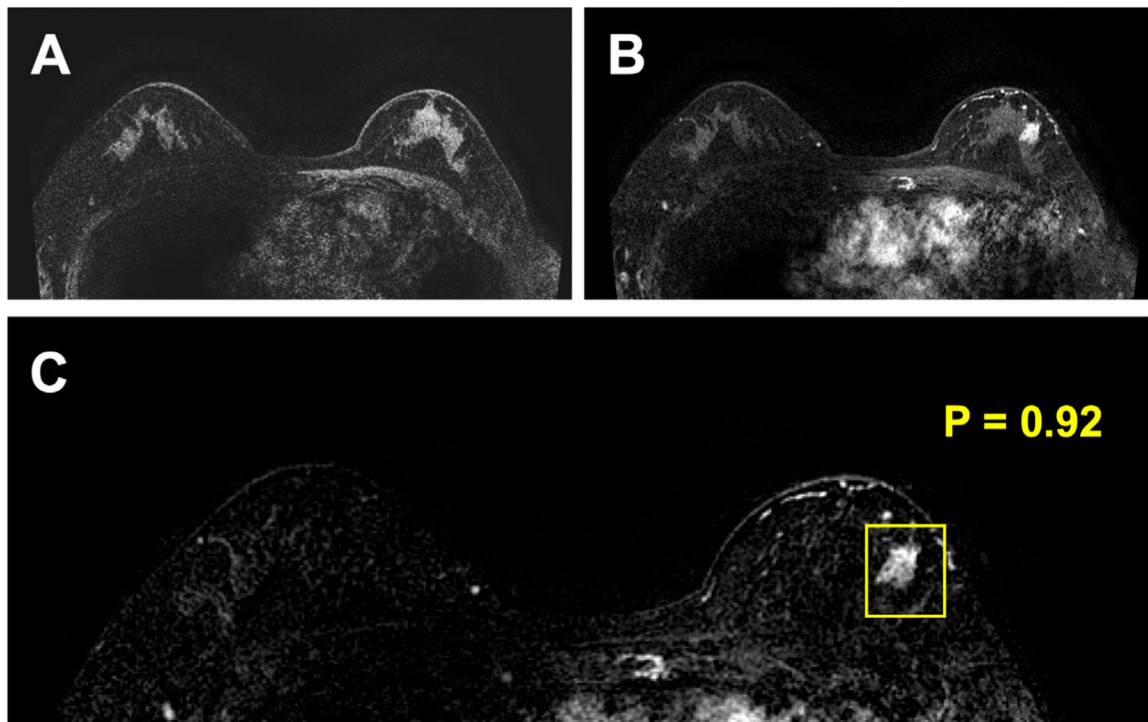
Figure 2.
Mask R-CNN architecture. A hybrid 3D-contracting and 2D-expanding fully convolutional feature-pyramid network is used as the backbone. The architecture incorporates the traditional 3×3 filters and the bottleneck 1×1–3×3–1×1 modules (left block). The number of input channel is 3, using the pre-contrast image and the subtraction images of the left and right breasts to utilize symmetry.
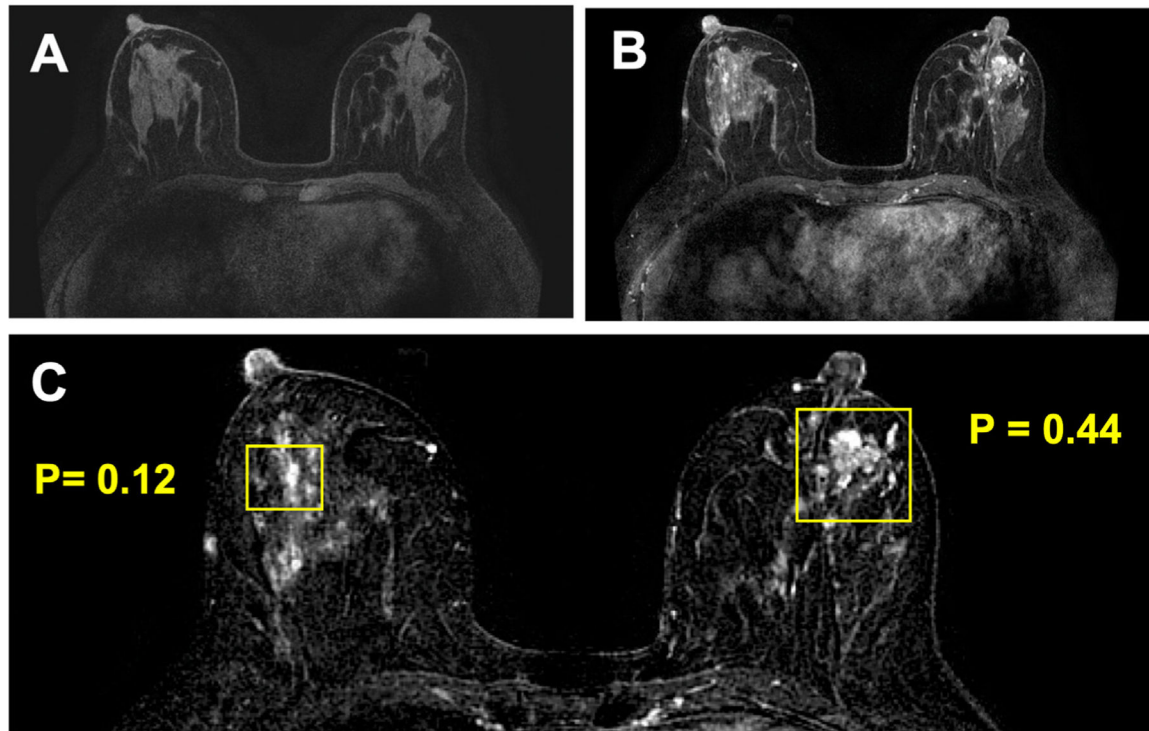
Figure 3.

The ResNet50 architecture consisted of 16 residual blocks. Each block contains one 1×1 convolutional layer, one 3×3 convolutional layer and one 1×1 convolutional layer. The detected bounding box from Mask R-CNN is used to crop the lesion on the three DCE parametric maps as the input, including Wash-in enhancement map (F2-F1)/F1, Mid-DCE enhancement map (F3-F1)/F1, and Wash-out Slope map (F6-F3)/F3. The output is a malignant probability, which is used to differentiate malignant when the probability is 0.5, and benign when the probability is <0.5.
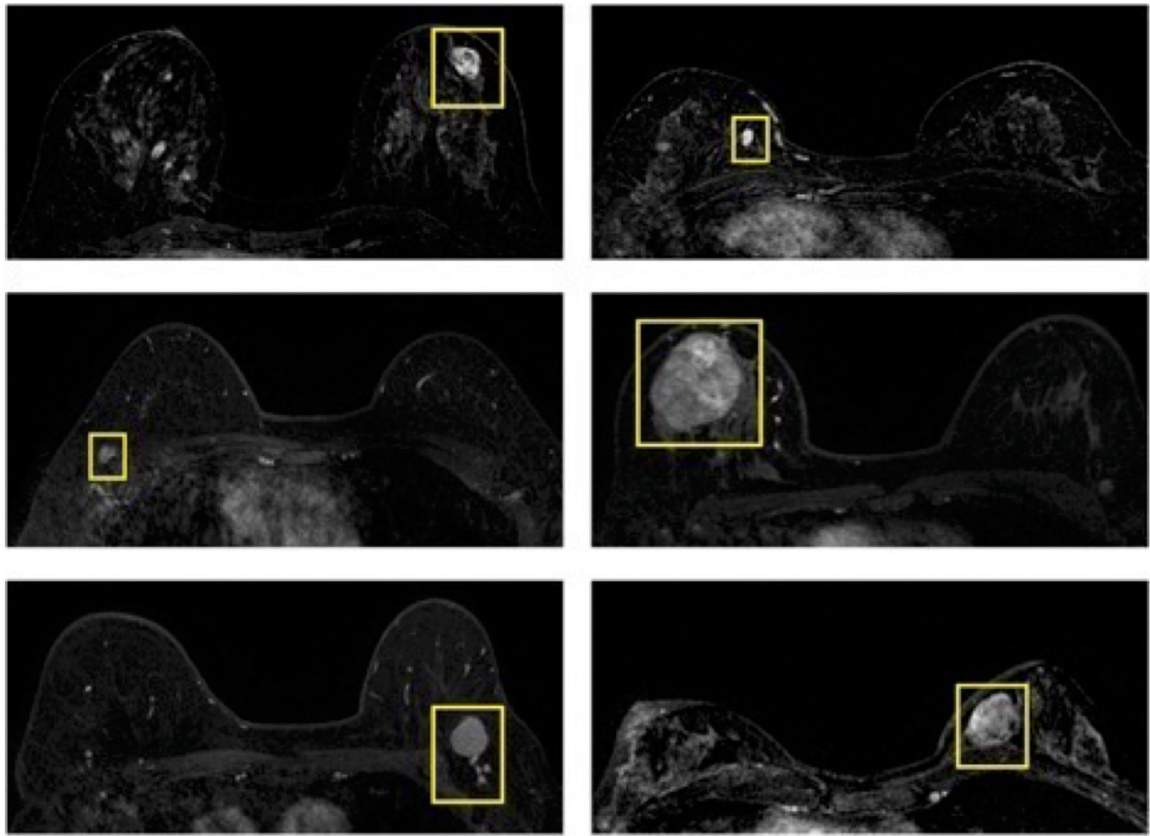
Figure 4.
True Positive (TP) case example from a 58-year-old woman with a small mass cancer (ductal carcinoma in situ). (A) Pre-contrast image acquired using fat-sat sequence; (B) Post-contrast image; (C) Tumor detection result searched by the Mask R-CNN algorithm, output as a box covering the suspicious lesion. The malignancy probability evaluated by the ResNet50 classification network is 0.92, suggesting a high likelihood of malignancy.
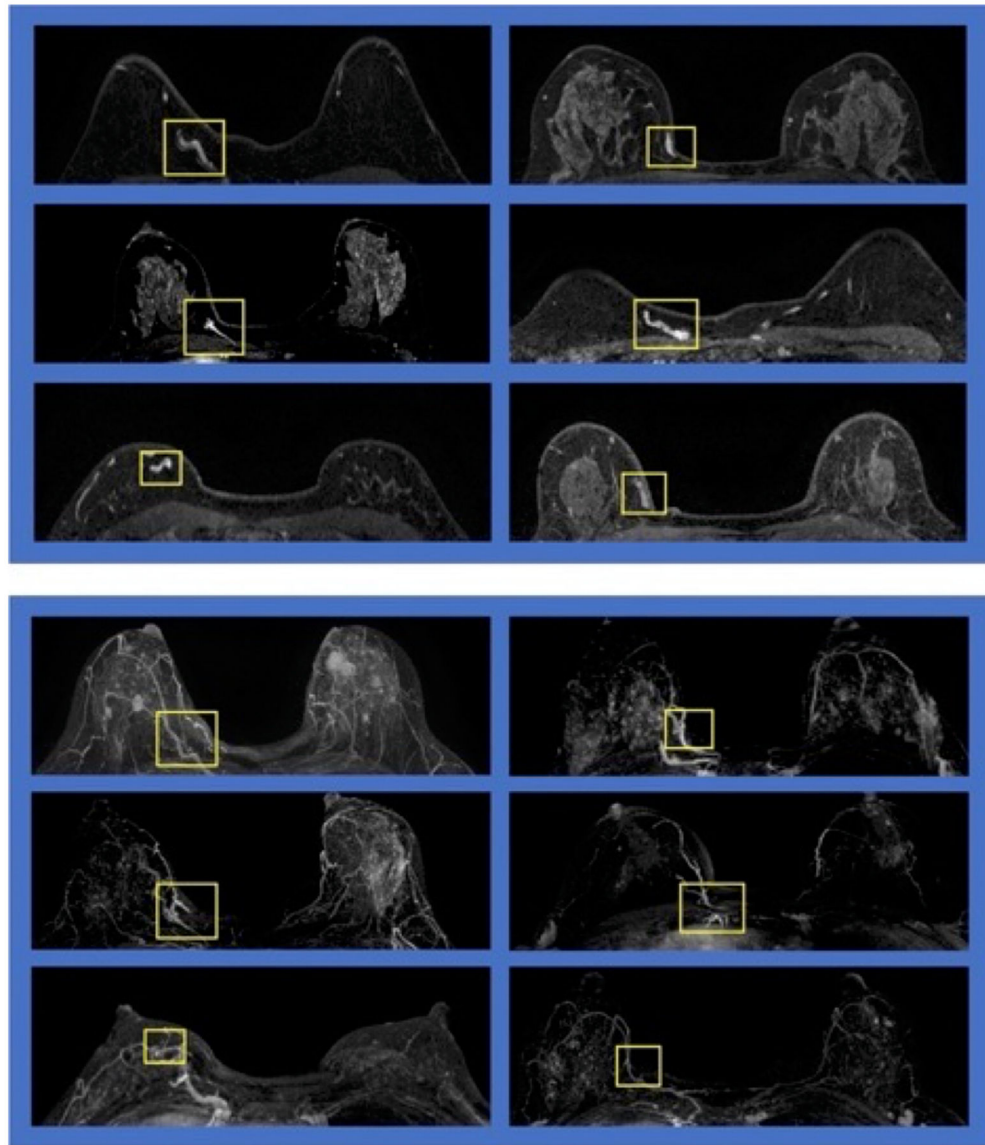
Figure 5.
Bilateral true negative (TN) example from a 44-year-old patient with a confirmed benign adenosis in the left breast. Extensive parenchymal enhancements are seen in both breasts. (A) Pre-contrast image; (B) Post-contrast image; (C) Tumor detection searched by Mask R-CNN. Two boxes are generated to identify two suspicious lesions, one in each breast. After evaluation by ResNet50, the left lesion shows a malignancy probability of 0.44, thus correctly diagnoses the pathologically confirmed adenosis as benign. The enhancements from the normal parenchymal tissues in the right breast has a low malignancy probability of 0.12.
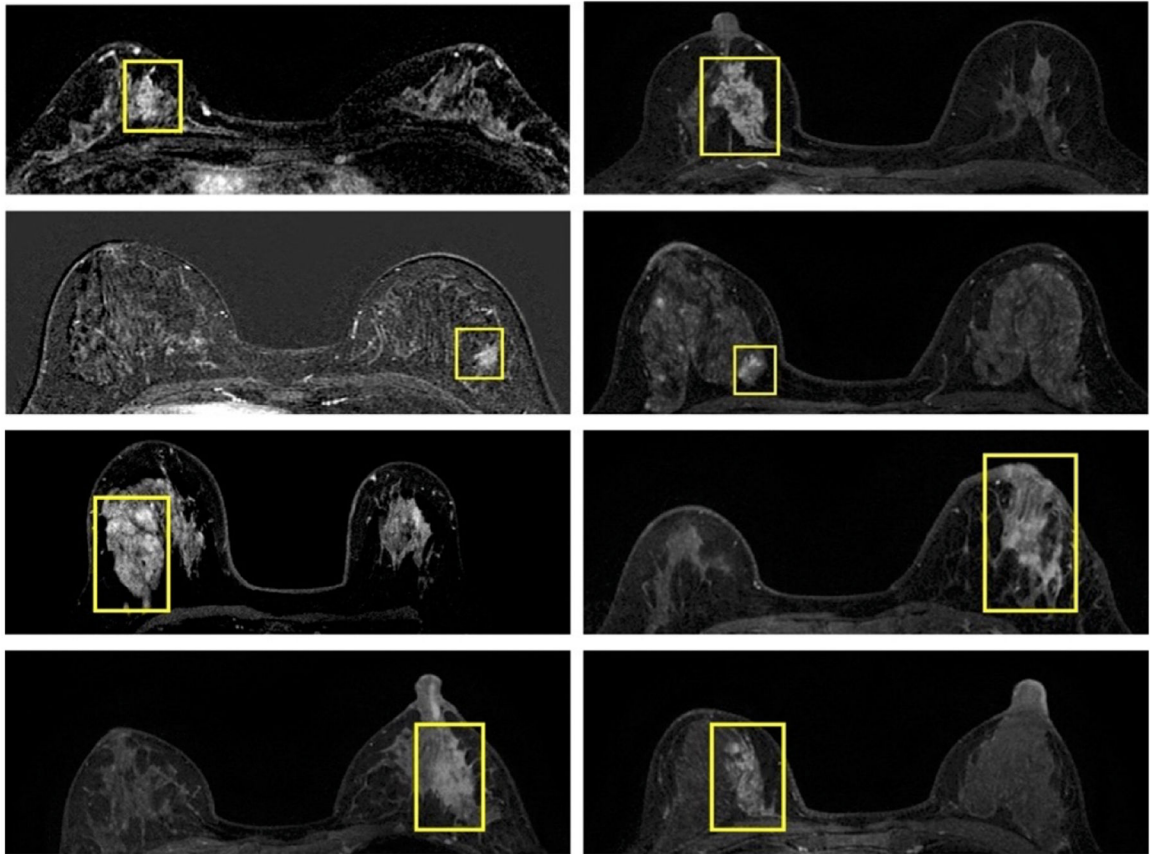
Figure 6.
Six false positive cases detected by Mask R-CNN followed by classification by ResNet50. They are from histologically confirmed benign lesions, showing strong enhancements with smooth boundary. Many other benign lesions can be eliminated by the ResNet50 classification network. Of a total of 73 benign lesions in the training dataset, only 16 are mis-diagnosed as cancer.

Figure 7.
Six false positive cases detected by Mask R-CNN followed by classification by Res-Net50. These detected lesions are coming from vessels, displaying strong linear enhancements. The top panel shows the identified areas on individual imaging slices, and the bottom panel shows the maximum intensity projection (MIP), used to confirm that the identified area is part of the vasculature.

Figure 8.
Eight false positive cases detected by Mask R-CNN followed by classification by ResNet50. These detected lesions are coming from breast parenchymal enhancements that do not present as symmetric between bilateral breasts. The parenchymal enhancements are usually irregular, larger, and show persistent enhancements in DCE sequence. Some of these may be related to benign processes such as fibrocystic changes or inflammation, but no biopsy-proven pathological diagnosis.

TABLE 1.

Pathological subtypes in malignant and benign groups in 2 datasets

PATHOLOGY	DATASET-1	DATASET-2
MALIGNANT	N= 103	N= 53
Invasive Ductal Cancer	75 (73%)	41 (77%)
Ductal Carcinoma *In-Situ*	21 (20%)	9 (17%)
Other Invasive Cancers	7 (7%)	3 (6%)
BENIGN	N= 73	N= 31
Adenosis	40 (55%)	16 (52%)
Fibroadenoma	16 (22%)	10 (32%)
Other Benign Lesions	17 (23%)	5 (16%)

TABLE 2.

Diagnostic performance of the initial detection by Mask R-CNN, and after classification by ResNet50

Dataset	Model	TP	FN	Sensitivity	*TN	*Specificity	Total FP	FP from Benign	FP from Vessels	FP from Parenchyma	PPV	FDR
Dataset-1	Mask R-CNN	101	2	98%	5	9%	179	48	33	98	36%	64%
	ResNet50	99	4	96%	37	70%	32	16	7	9	76%	24%
Dataset-2	Mask R-CNN	49	4	92%	4	13%	121	27	23	71	29%	71%
	ResNet50	43	10	81%	25	81%	28	6	8	14	61%	39%

*
Specificity is referring to the diagnosis of histologically confirmed benign lesions.TP: True Positive; FN: False Negative;

*
TN: True Negative of confirmed benign lesions; FP: False Positive; Sensitivity = TP/(TP+FN); Specificity = TN/(TN+FP from Begin); PPV: Positive Predictive Value = TP/(TP+Total FP), FDR: False Detection Rate = FP/ (FP+TP).