

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Fully automated detection of breast cancer in screening MRI using convolutional neural networks

Mehmet Ufuk Dalmış
Suzan Vreemann
Thijs Kooi
Ritse M. Mann
Nico Karssemeijer
Albert Gubern-Mérida

Fully automated detection of breast cancer in screening MRI using convolutional neural networks

Mehmet Ufuk Dalmış,* Suzan Vreemann, Thijs Kooi, Ritse M. Mann, Nico Karssemeijer, and Albert Gubern-Mérida

Radboud University Medical Center (RadboudUMC), Diagnostic Image Analysis Group (DIAG) Nijmegen, The Netherlands

Abstract. Current computer-aided detection (CADe) systems for contrast-enhanced breast MRI rely on both spatial information obtained from the early-phase and temporal information obtained from the late-phase of the contrast enhancement. However, late-phase information might not be available in a screening setting, such as in abbreviated MRI protocols, where acquisition is limited to early-phase scans. We used deep learning to develop a CADe system that exploits the spatial information obtained from the early-phase scans. This system uses three-dimensional (3-D) morphological information in the candidate locations and the symmetry information arising from the enhancement differences of the two breasts. We compared the proposed system to a previously developed system, which uses the full dynamic breast MRI protocol. For training and testing, we used 385 MRI scans, containing 161 malignant lesions. Performance was measured by averaging the sensitivity values between 1/8—eight false positives. In our experiments, the proposed system obtained a significantly ($p = 0.008$) higher average sensitivity (0.6429 ± 0.0537) compared with that of the previous CADe system (0.5325 ± 0.0547). In conclusion, we developed a CADe system that is able to exploit the spatial information obtained from the early-phase scans and can be used in screening programs where abbreviated MRI protocols are used. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.5.1.014502]

Keywords: breast MRI; lesion detection; deep learning; computer-aided detection; screening.

Paper 17147R received May 24, 2017; accepted for publication Dec. 18, 2017; published online Jan. 11, 2018.

1 Introduction

Breast magnetic resonance imaging (MRI) is known for its high sensitivity in detecting breast lesions. It has been shown that lesions that are occult in mammography and ultrasonography can be detected in breast MRI.¹ In a typical breast MRI acquisition protocol, after an initial T1-weighted (T1w) MRI scan is obtained, a contrast agent is administered to the patient to enhance lesions, and, subsequently, several postcontrast T1w MRI scans are obtained. Lesions become visible in the subtraction volume obtained from precontrast and the first postcontrast volumes, which are referred to as early-phase scans. The additional T1w scans obtained after the first postcontrast MRI are used for evaluating contrast enhancement dynamics of a lesion in the late phase, which provides adjunct information for distinguishing the benign structures from the malignant ones.²

Despite the higher sensitivity of breast MRI, mammography remains as the standard modality for general screening of women for breast cancer since high cost of breast MRI limits its widespread use. One of the cost-increasing factors is the acquisition of several scans obtained for a single breast MRI study. To decrease the cost and be able to facilitate the application of this imaging modality in screening, abbreviated breast MRI protocols have been suggested,^{3,4} where the late-phase T1w acquisitions of the full DCE-MRI protocol are not obtained to reduce the protocol time. In a simple abbreviated protocol, evaluation of an enhancing structure is based only on morphological information obtained from the early-phase scans. Additionally, a quick reading protocol was suggested for the abbreviated protocols

since interpretation of high-dimensional images is time consuming for radiologists. In this quick reading protocol, maximum-intensity projection (MIP) images of the first postcontrast subtraction volumes are used. The radiologist initially decides on the presence of suspicious regions based on the MIP image. Then, in case of a positive decision, the three-dimensional (3-D) subtraction volume is also inspected for a final conclusion. This reading method reduces the reading time from a reported average of 4.7 min⁵ to <1 min.^{3,4} However, this quick reading protocol may increase reading errors. Such reading errors are not uncommon even when a full diagnostic breast MRI protocol is used. Several studies^{6–8} showed that up to 56% of the cancer lesions detected in MRI could have been detected in earlier MRI scans, but they were misinterpreted or overlooked by the readers. Studies on abbreviated protocols indicate that reading errors may occur more often when a quick reading protocol is used, as it was found that some lesions were missed due to the use of MIP images as a first step in the reading workflow.^{3,4}

Computer-aided detection (CADe) systems^{9–12} have been developed to aid radiologists reading breast MRI. These systems have the potential to reduce reading time and prevent reading errors since they are able to detect lesions that were misinterpreted or overlooked by radiologists during breast MRI screening.¹³ However, current breast MRI CADe algorithms rely on the full breast MRI protocol including the temporal information from the late-phase scans, in addition to the morphological information obtained from the early-phase scans. This is an important limitation since it limits their applicability on screening programs using an abbreviated breast MRI protocol where

*Address all correspondence to: Mehmet Ufuk Dalmış, E-mail: mehmet.dalmis@radboudumc.nl

only early-phase scans are available and assessment is based only on morphology. Moreover, it is questionable whether these existing CAde systems are able to fully exploit the morphological information. In a previous study,¹⁴ where 395 lesions were included from 325 patients, it was shown that the performance of the computer-aided diagnosis system mainly relied on the dynamic features, whereas in clinical assessment, morphology is the most vital information and dynamic information is auxiliary. Automatic evaluation of lesion morphology in a conventional CAde system is difficult since it requires design of specific features to be extracted from images. Furthermore, designing such features is known to be the most difficult part and main performance limiting factor of conventional computer vision systems. Recently, popular deep-learning methods tackle this difficulty by learning such features automatically based on examples, instead of using human-engineered features, often using convolutional neural networks (CNN). This approach turned out to be a ground-breaking success in computer vision compared with the traditional methods.^{5,15} In accordance to the mentioned advances in the computer vision field, medical image analysis has also extensively benefited from the deep neural networks¹⁶ in various domains, such as ophthalmology,¹⁷ dermatology,¹⁸ histopathology,¹⁹ brain,²⁰ and prostate²¹ and breast²² MRI. Therefore, in this study, we focused on developing a CAde system, which relies on exploiting the spatial information available in the early-phase MRI volumes using a deep-learning approach. For this purpose, we developed a fully automated lesion detection pipeline consisting of several deep learning components. To our knowledge, this is the first application of deep learning to fully automated lesion detection in breast MRI.

We compared the proposed CAde system with a previously developed CAde system that uses the full dynamic breast MRI protocol (early and late-phase scans).^{11,13} We used the same training and test sets as in the study by Gubern-Mérida et al.,¹³ which allowed us to make a direct comparison with the results presented in our previous work. The test set included lesions visible in prior examinations that were assessed as negative during screening practice. We report the performance of the proposed CAde system individually for screening-detected lesions and for lesions that were visible but missed in the prior MRI examinations, as well as for the overall dataset.

2 Materials and Methods

2.1 Dataset

The training dataset was composed of breast MRI scans of 201 women, who underwent MRI for various reasons, including

screening and preoperative staging. About 87 of the MRI scans contained a total of 95 visible malignant lesions. The average effective radius of the lesions was 10.8 mm with a standard deviation of 6.2 mm and a range of 2.5 to 29.8 mm. The remaining 114 MRIs were considered normal since they were scored as either BI-RADS 1 ($n = 87$) or BI-RADS 2 ($n = 27$) by the radiologists. For the women with normal MRI scans, at least 2 years of follow-up were available with no signs of breast cancer, and no previous history of breast cancer or breast surgery was reported.

The testing dataset was composed of MRI scans of women participating in a high-risk screening program between 2003 and 2014. Of the 160 women in the test set, 120 had normal MRI scans with no signs for malignancy. These women had no history of breast cancer or surgery and had at least 2 years of follow-up with no signs of breast cancer (BI-RADS 1 or 2). The remaining 40 women were diagnosed with breast cancer detected on MRI, with a total of 42 malignant lesions. These women also had MRI examinations performed one year earlier, which were classified as negative (BI-RADS 1 or 2). After detection of the cancers, these prior-negative MRI scans were re-evaluated retrospectively by two radiologists in consensus. In 24 of these prior scans, lesions were detected retrospectively, whereas lesions were not visible in the remaining 16 scans. Of the 24 lesions (of 24 scans) that were detected in the prior-negative scans, 11 lesions were classified as “visible” (BI-RADS 4/5) and 13 lesions were classified as “minimal sign” (BI-RADS 2/3) by the two radiologists in consensus. The visible lesions are referred to as “prior-visible” lesions, whereas the lesions with minimal signs are referred to as “prior-minimal sign.” In total, we had 66 lesions in the testing dataset (42 screening-detected, 11 prior-visible, and 13 prior-minimal sign). The lesions in this dataset were smaller than the ones in the training dataset since these lesions were detected in a high-risk screening program. The average effective radius of the lesions in this dataset was 4.8 mm with a standard deviation of 2.5 mm and a range of 2.0 to 15.8 mm. In Fig. 1, we provided a few examples to these lesions.

The MRI scans included in this study were obtained from 1.5 or 3 Tesla Siemens scanners using a dedicated bilateral breast coil. A gradient-echo dynamic sequence was performed to obtain a T1w MRI before the administration of a contrast-agent [Gd-DOTA (Dotarem, Guerbet, France), at a dose ranging from 0.1 to 0.2 mmol/kg]. Within 2 min after the administration of the contrast agent, the first postcontrast T1w scan was obtained, which was followed by three to five additional T1w scans. Acquisitions were either in transversal or coronal planes with pixel spacing in 0.664- to 1.5-mm range and slice thickness in 1- to 1.5-mm range. Other MRI acquisition parameters were

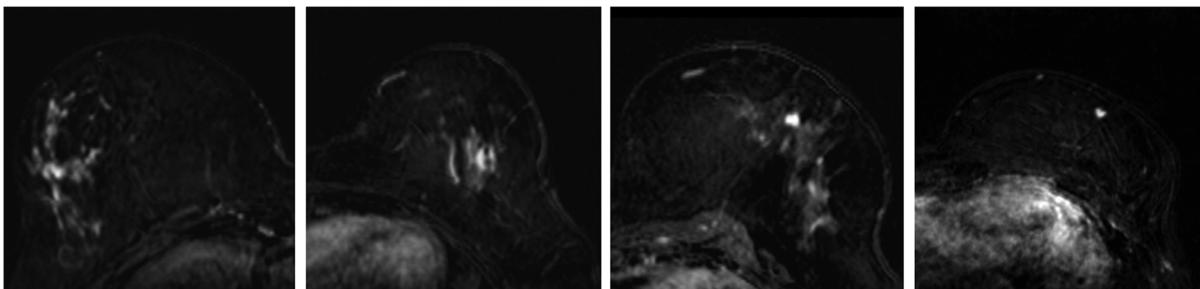


Fig. 1 A few examples of lesions that were detected in the screening program.

1.71 to 4.76 ms for echo time, 4.56 to 8.41 ms for repetition time, and 10 deg to 25 deg for flip angle.

2.2 Annotation

All lesions were annotated by radiologists in an in-house developed dedicated breast MRI workstation, which includes a semi-automated tool “smart opening”^{23,24} for lesion segmentation. With this tool, a 3-D lesion segmentation is obtained after the annotator places a seed-point at the center of a lesion. Annotators were able to add more seed-points when the result was not satisfactory especially for the large lesions. Each lesion was classified by the annotators as mass or nonmass. Motion-corrected subtraction volumes were used in this process.¹¹

2.3 Preprocessing

For each MRI examination, postcontrast T1w volumes were registered to the precontrast T1w volume to correct for motion using the Elastix toolbox.²⁵ Subsequently, the subtraction volume was obtained by subtracting the precontrast image from the motion-corrected first postcontrast image. The relative enhancement (RE) volume was also obtained, which is computed by normalizing the subtraction intensities relative to the precontrast intensities using the following equation:

$$RE = \frac{I_1 - I_0}{I_0}, \quad (1)$$

where I_0 and I_1 are the intensity values in precontrast and motion-corrected first postcontrast images, respectively.

2.4 Automated Lesion Detection

The automated lesion detection system developed in this study, as shown in Fig. 2, uses two images as inputs: the precontrast T1w volume and the RE volume. Given these inputs, the lesions are detected in three steps. Initially, the breast is segmented based on the precontrast T1w image, and subsequently, this segmentation mask is applied to the RE image. At the second step, a candidate detection algorithm uses the RE image to search for possible candidate locations within the segmented breast. Here, we define the term “candidate” as a voxel or location, which is expected to represent a lesion in its local neighborhood in the breast MRI volume. At the final step, the candidates are classified to reduce the false positives of the previous stage.

2.4.1 Breast segmentation

A fully automated breast segmentation method based on deep learning²² was used in this study to segment the breasts in

MRI volumes. This method is based on a two-dimensional (2-D) U-net architecture.²⁶ U-net is a fully convolutional network, which produces “dense prediction.” In the other words, for each pixel or voxel in the input image, U-net generates a likelihood value, which is a useful property for segmentation or detection problems. In a U-net, this is achieved using the de-convolutional part of the network that comes after the convolutional part, where the output of the convolutional part is up-sampled. Each axial slice is provided individually to this algorithm to generate the corresponding likelihood map, where a voxel value in this map indicates the likelihood of the voxel to belong to the breast. The final segmentation for an axial slice is obtained by thresholding the likelihood values at 0.5. The 3-D segmentation of the breasts is obtained by combining these 2-D segmentation slices into a 3-D volume.

2.4.2 Candidate detection

To detect candidate locations, we followed a similar approach as the one used to segment the breasts. We trained a two-level version of the same U-net that was used in the breast segmentation algorithm, using axial slices to identify the voxels that belong to a lesion. For this purpose, the lesion segmentations manually generated by the radiologists were used as the ground-truth. About 20% of the training dataset was separated at the patient level to be used as a validation set. The U-net was trained in batches of 40 randomly selected axial slices: 20 slices containing a segmented lesion and 20 slices without a segmented lesion. The slices for each batch were randomly selected from the entire dataset, where the total number of slices containing a segmented lesion was 4106. The loss function used during training was the log of the absolute value of the differences between the target and output likelihood values, averaged over the breast region of the given slice. We used Glorot-uniform initialization²⁷ for weights of the U-net and RMSProp²⁸ for gradient-descent optimization. The initial learning rate was set as 0.001 and it was divided by 10 at each 1000 iteration. We continued training until the loss in the validation set was stable. The resulting trained U-net, given an MRI axial slice, outputs a likelihood map for each voxel of the given MRI slice to belong to a lesion. An example result for an MRI slice is given in Fig. 3. Subsequently, we obtained a likelihood volume for each MRI, combining these 2-D likelihood maps. The final candidates were obtained by applying a local maxima algorithm¹¹ on the likelihood volumes.

2.4.3 Candidate classification

In the first candidate detection stage, a set of candidates was obtained based on 2-D shapes and patterns since the U-net

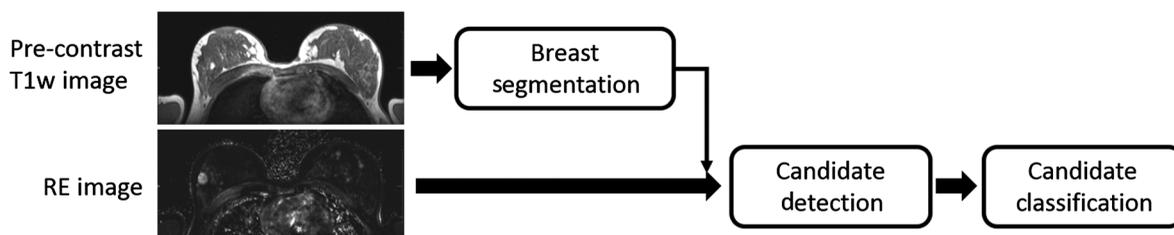


Fig. 2 Pipeline for the proposed CADe system. It uses two inputs: the precontrast volume to be used in breast segmentation, and the registered first postcontrast RE volume. Lesion candidates which are detected in the segmented breast region are classified in the last step.

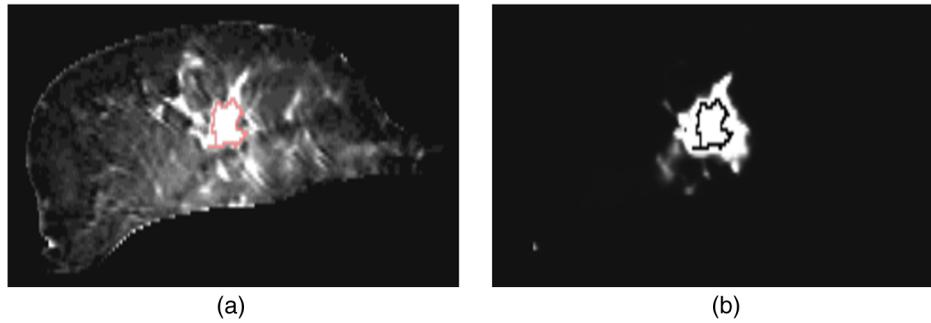


Fig. 3 U-net candidate detection example for an MRI slice. (a) The corresponding slice in the breast-segmented RE volume. (b) Corresponds to the lesion likelihood map for the same slice, output by the candidate detection U-net. Contours on both images represent the segmented lesion for this slice.

was applied in a slice-by-slice basis. In the second and final step, the likelihood for each candidate is further refined to reduce false positives. For this purpose, we employed a 3-D CNN that uses two types of information available in the RE volumes: 3-D spatial (morphological) information in the local region around the candidate, and the information arising from the asymmetry between the enhancements of the two breasts.

In previous work,^{29–31} it has been shown that lesion detection accuracy can be improved by taking symmetry information of the contralateral breast into account. In Refs. 30 and 31, this symmetry information was incorporated in a patch-based scheme. We followed a similar approach to exploit the symmetry information in a patch-based fashion. For each candidate location in a breast, we computed the corresponding location in the contralateral breast, using the breast masks obtained in the breast segmentation stage. As shown in Fig. 4, we initially computed the center of gravity (CoG) for each breast. We considered the CoG of a breast as the origin of a 3-D coordinate system for the given breast. The coordinate systems for right and left breasts were mirrored with respect to the median plane. The location of the candidate in this coordinate system was applied to the coordinate system of the contralateral breast to find the corresponding location of the candidate.

For each candidate, a patch of $39 \times 39 \times 39$ voxels around the candidate location and a patch of the same size around the corresponding location in the contralateral breast were extracted. These two patches were input to a 3-D CNN, which was trained to classify positive and negative candidates. The selection of the patch size was based on the consideration to include sufficient contextual information while keeping the

computational complexity at a reasonable level. The use of an odd number was for practical reasons, to include regions around the candidate in a symmetric way. The architecture of the CNN used for this purpose is shown in Fig. 5. For each of the two inputs, we used two convolutional layers of $5 \times 5 \times 5$ with rectified linear unit (ReLU) transfer functions, each followed by a maxpooling layer of size $2 \times 2 \times 2$. This was followed by a fully connected (dense) layer with a ReLU transfer function. These two streams corresponding to the two inputs shared the same weights, and they were combined to be input to a second dense layer with a ReLU transfer function. A final dense layer with softmax transfer function was used to compute the final likelihood values.

We used a validation set to determine the hyperparameters of the network and the training process and to monitor the performance during training. For this purpose, 20% of the training dataset was randomly selected at the case level. Area under the ROC curve (AUC) was used to measure the performance in this set. We used Glorot-uniform initialization²⁷ for weights of the network and RMSProp²⁸ for gradient-descent optimization. Drop-out was applied to the last dense layer with a rate of 0.85 and a starting learning rate of 0.001 was selected. At each batch we used 32 positive and 32 negative candidates, which were randomly selected from the entire dataset. As the training continued, the learning rate was dropped by half when the performance in the validation set did not improve further for 50 epochs, where each epoch consisted of eight batches. The training was stopped when the performance in the validation set was not improved in the last 200 epochs. The final model was selected as the model resulting in the highest area under the ROC curve (AUC) on

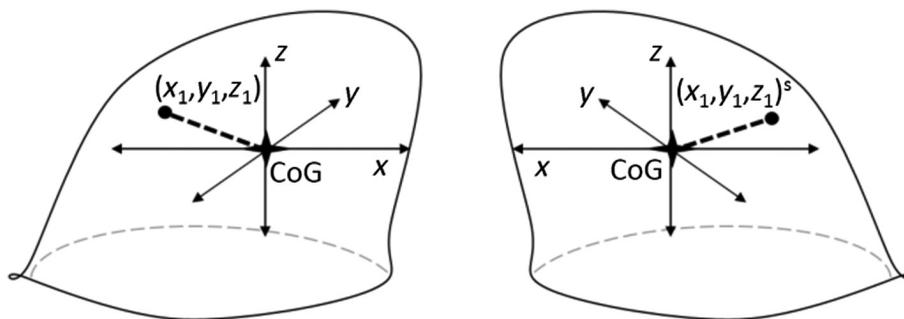


Fig. 4 Given a location (x_1, y_1, z_1) in the coordinate system of one breast, the corresponding location $(x_1, y_1, z_1)^s$ in the contralateral breast was identified. Each coordinate system had the origins at CoG of the breast and they were mirrored to each other along the medial plane.

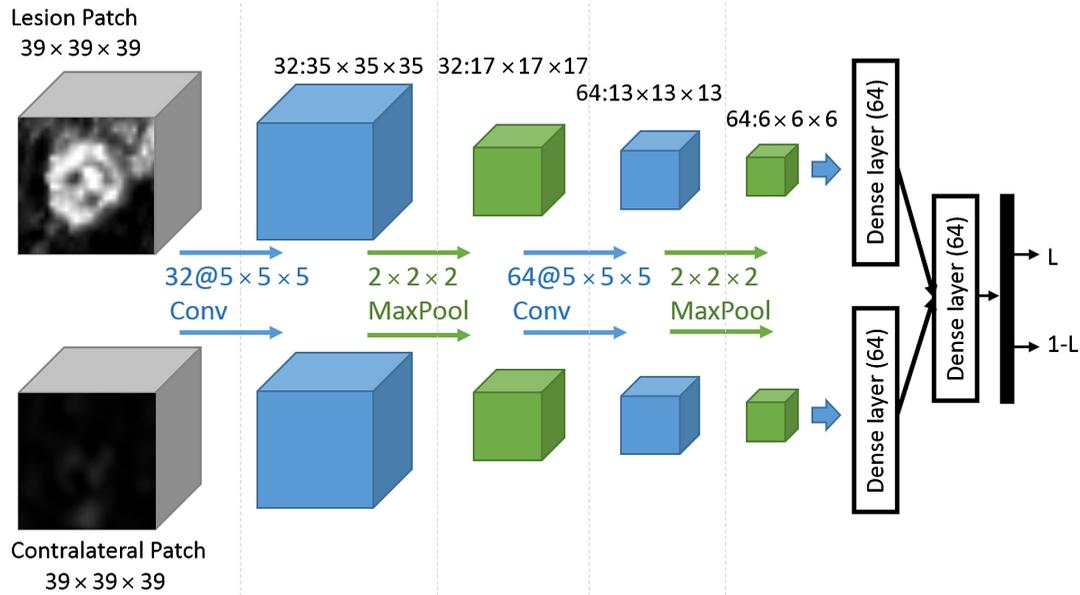


Fig. 5 The CNN used in the study. There are one convolutional, one max-pooling, and one dense layer for each input, where the weights are shared. After the two streams are concatenated, an additional dense layer and a final softmax layer are used to obtain lesion likelihood values (L) for each candidate.

the validation set. For data augmentation, we applied random translations (maximum five voxels in each direction), rotations and mirroring to the input patches.

2.5 Experiments and Evaluation Method

2.5.1 Experiments

We compared the proposed CAde system to an existing CAde system.¹¹ This system uses the full dynamic breast MRI protocol: the precontrast and all registered postcontract images. It consists of the following steps: breast segmentation, candidate detection, and false positive reduction. For breast segmentation, an atlas-based method was used. For the first stage detection, a likelihood map of the MRI was produced using a voxel classifier, which was trained using several RE and blob feature maps. Using this likelihood map, candidates were extracted using the local maxima algorithm, which was also used in the presented study. Finally, false positives were reduced in a second-stage classifier, which used both morphological and contrast-dynamics information. Note that this system was trained and tested on the same datasets as the ones used to evaluate the presented pipeline.

To study the effect of symmetry information on the performance of the proposed CAde system, we trained another 3-D CNN that only uses the candidate patch coming from the suspicious breast as an input. The input for the contralateral patch and its corresponding stream was removed from the CNN and training was performed using the same hyperparameters. We refer to this system as CAD-WoS, where WoS stands for “without symmetry information.”

We reported the performance of the proposed system for different lesion types (mass-like and nonmass-like lesions) and for different lesion subsets (screening-detected, prior-visible, and prior-minimal sign lesions).

We used the Titan X graphical processing unit (GPU) of NVIDIA[®] for deep-learning experiments.

2.5.2 Evaluation method

A candidate was considered as a true positive when its location is within the manually segmented volume of a lesion, and it was considered as a false-positive otherwise. When multiple candidates hit the segmented volume of a lesion, the candidate with the highest likelihood was chosen. We used free-response operating characteristic (FROC) analysis to assess the performance of the evaluated systems. For each threshold level on the final likelihood values of the candidates, the average number of false positives in normal images in the test set was computed. To obtain a final performance metric for the system, we used the computation performance metric (CPM),³² where sensitivity values at 1/8, 1/4, 1/2, 1, 2, 4, and 8 false positives per scan were averaged.^{33–37}

Statistical comparison between two FROC curves was performed using the bootstrapping method.³⁸ We sampled cases 1000 times with replacement and constructed the FROC curves for the two systems based on these samples. For each bootstrapped curve, we computed the difference between CPM values (Δ CPM). The p value for a statistical comparison was defined as the number of the negative or zero valued Δ CPM’s divided by the number of samples, 1000. We used Bonferroni correction for three comparisons. The difference between two CPM values was considered as significant when the p value was $<0.05/3$.

3 Results

The FROC plots for the proposed CAde system with and without symmetry information, and for the previous CAde system are given in Fig. 6. The CPM value obtained by the proposed CAde system was 0.6429 ± 0.0537 , which was significantly higher than the CPM value of 0.5325 ± 0.0547 obtained by the previous CAde system ($p = 0.008$). The CPM value obtained by the CAD-WoS was 0.5804 ± 0.0572 . The differences in CPM values between the previous CAde system and the CAD-WoS system, and between the proposed CAde system

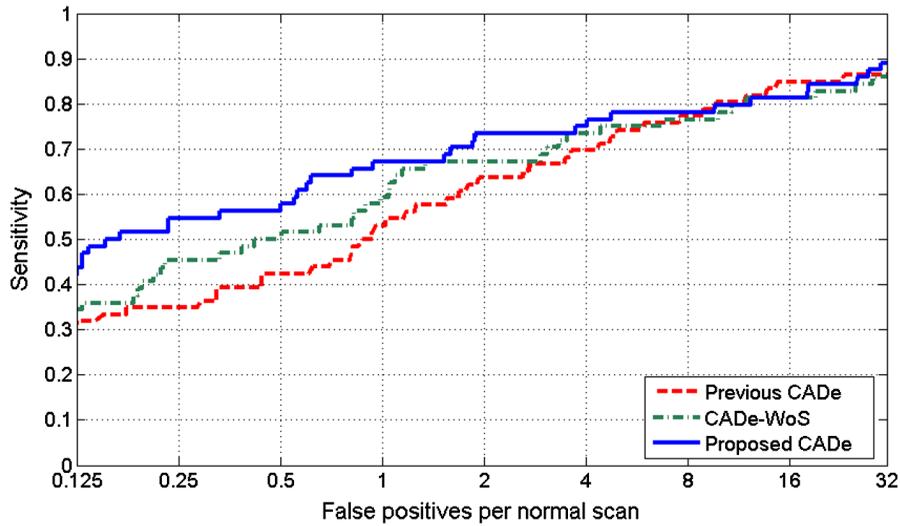


Fig. 6 FROC plots for the proposed CADE system, CADe-WoS system, and the previous CADE system.

and CADe-WoS system were not statistically significant ($p > 0.05/3$).

The FROC plots for different lesion subsets are given in Fig. 7. Corresponding CPM values are given in Table 1.

CPM values increased in all lesion subsets compared with the CPM values obtained by the previous CADE system. We observed the most remarkable performance improvement compared with the previous CADE system in the prior-visible

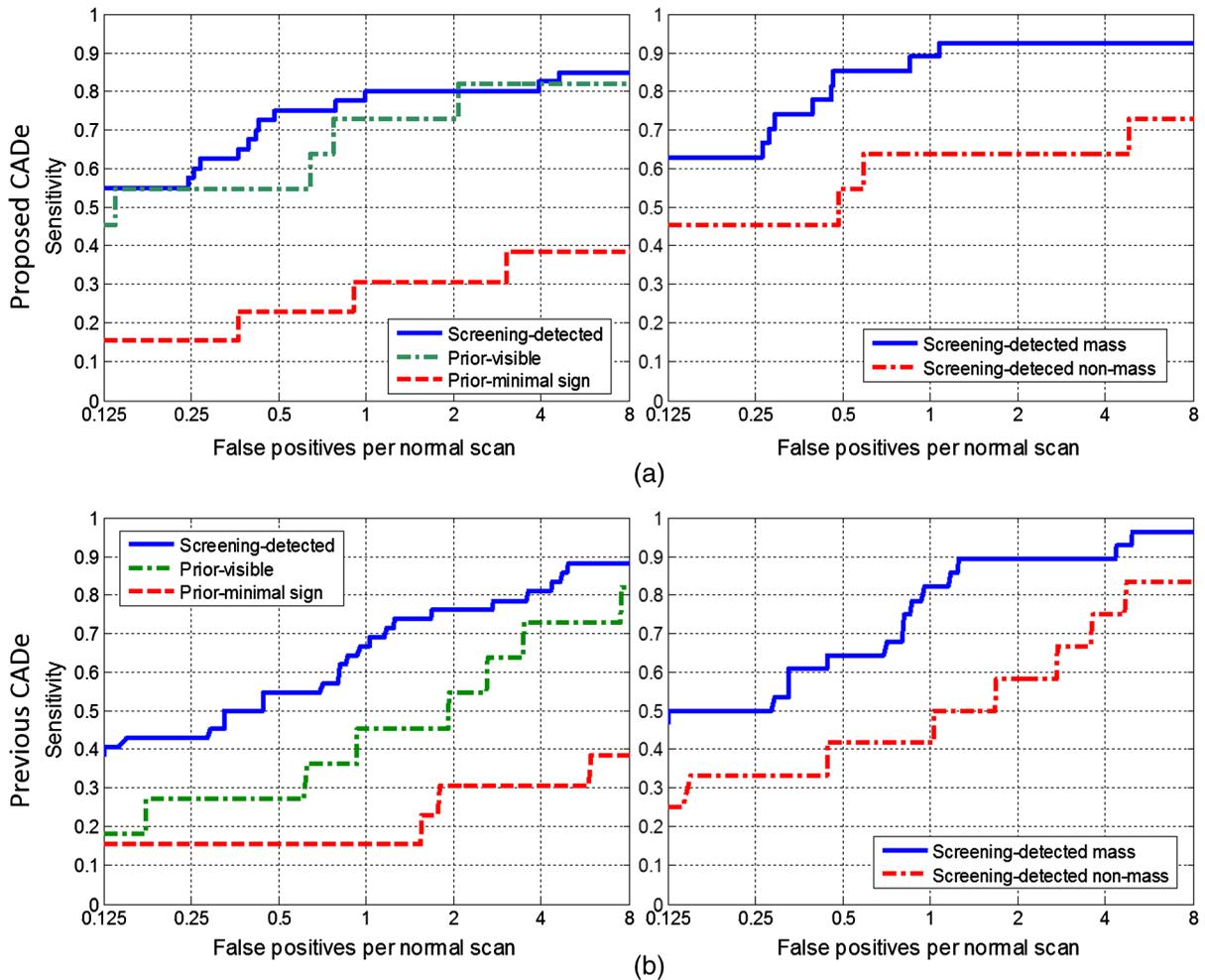


Fig. 7 Comparison of performances of the two CADE systems in different lesion subsets: (a) screening-detected, prior-visible, and prior-minimal sign lesions and (b) mass and nonmass lesions.

Table 1 CPM values for different lesion types.

	Previous CADe	Proposed CADe
All lesions	0.5325	0.6429
Screening-detected	0.6803	0.7357
Screening-detected mass	0.7685	0.8253
Screening-detected nonmass	0.4935	0.5844
Prior visible	0.4675	0.6623
Prior-minimal sign	0.2308	0.2747

lesions (CPM increase from 0.4675 to 0.6623) and in the nonmass lesions (CPM increase from 0.4935 to 0.5844).

4 Discussion

In this study, we developed a fully automated CADe system for breast MRI using deep learning. Our purpose was to exploit the spatial information obtained in the early-phase as much as possible, without using the temporal information from the late-phase of the contrast enhancement. We followed a deep-learning approach since deep learning has advantages over the conventional computer vision methods, such as using automatically learned features rather than handcrafted features to evaluate spatial information. The proposed CADe system obtained a CPM score of 0.6429, which was significantly ($p = 0.008$) higher than the CPM score of 0.5325 obtained by the previous CADe system that used the full dynamic breast MRI protocol.

The developed CADe system uses symmetry information arising from the differences between the contrast enhancements of the two breasts of the same woman, in addition to the 3-D morphological information in the candidate regions. A previous study that used symmetry information for automatically detecting lesions in breast MRI had shown its benefit for DCIS lesions.²⁹ Different than the referred study, we used symmetry information in a deep-learning scheme using a similar method as employed for mammography images,³⁰ albeit in 3-D. Our results showed that symmetry information contributes to improve the overall performance of the CADe system (CPM values of 0.6429 and 0.5804 with and without using symmetry information, respectively). This was expected since symmetry information is also used by the radiologists to assess lesions and it is stated in the guidelines.² The performance of the CADe system without symmetry information was still higher than the performance of the conventional CADe system that uses the full dynamic breast MRI protocol.

The CNN architecture we used in the candidate classification stage was not very deep, including two convolutional layers with $5 \times 5 \times 5$ sized filters and three dense layers. This is similar to the LeNet5,³⁹ which was developed for handwritten digit recognition, except that we used a 3-D CNN. In our experiments, we found that it was essential to limit the complexity of the model to prevent overfitting, even when a high drop-out rate was used for regularization. This may be related to the limited amount of data used for training. Deeper networks are useful in natural images as they include complex hierarchical relations that require many layers of representations. Although this may not be the case for medical images, in particular for MRI scans, more complex and

deeper models may still be useful in the presence of a large amount of data.

To reflect a real screening situation, we evaluated the performance of the developed CADe system on a test set including cancer and normal cases from woman participating in a high-risk screening program with MRI. However, the training set did not include screen-detected cancers because we did not have them available at the time of the study. Occurrence of malignant lesions in screening breast MRI scans is much less frequent compared with the diagnostic MRI scans. The test set included lesions that were visible or had minimal signs in prior scans, which allowed us to test the performance of the system for lesions that were overlooked or misinterpreted by the radiologists. The most remarkable increase in detection performance was observed in lesions visible in prior MRI scans (see Fig. 7 and Table 1). The proposed CADe system was able to detect 60% more lesions compared with the conventional CADe system at the threshold equivalent to 1 false positive per scan. This suggests that the proposed CADe system might be more useful to prevent reading errors.

We also tested the performance of the proposed CADe system individually for mass-like and nonmass-like lesions (see Fig. 7 and Table 1) within the screening-detected subset of the test set. We observed a performance increase in the detection of both lesion types. However, the detection performance for nonmass lesions was still lower than that for the mass lesions. More training data of nonmass lesions and the development of dedicated algorithms may be needed to increase detection performance for such lesions.

Our study has some limitations. We did not evaluate how the addition of contrast dynamic information may improve the performance of the presented CADe system. Although contrast dynamic information from the late-phase may not be available in a screening setting, fast MRI acquisition protocols make it possible to obtain dynamic contrast uptake information in the early-phase of the acquisition, which does not cause an increase in the duration of the breast MRI. It should be investigated in a further study whether the addition of such early-phase dynamic information can increase the detection performance of a CADe system. Additionally, we assumed that the two breasts have the same (mirrored) shapes to identify corresponding locations in the contralateral breasts. Although this assumption holds for most of the cases, some patients may have asymmetric breast shapes. In the future, we will investigate more advanced registration methods that take into account the differences in shapes between the right and left breasts of a given woman.

In conclusion, we developed a CADe system for breast MRI that uses only the early-phase of the acquisition without using dynamic information from the late-phase T1w acquisitions. To fully exploit the spatial information obtained in the early-phase, we used 3-D morphology of the candidate regions and symmetry between the enhancements of two breasts of the patient, using a deep-learning approach. The proposed CADe system significantly outperformed a conventional CADe system which uses the full dynamic breast MRI protocol. The developed CADe system can be used in abbreviated MRI protocols that have been suggested for MRI screening programs.

Disclosures

Nico Karssemeijer is cofounder and shareholder of Volpara Solutions Ltd. (Wellington, New Zealand), cofounder and shareholder of QView Medical Inc. (Los Altos, California) and

cofounder and shareholder of QView Medical ScreenPoint Medical BV (Nijmegen, The Netherlands).

Acknowledgments

This work was funded by the European 7th Framework Program grant VPH-PRISM (FP7-ICT-2011-9, 601040). We gratefully acknowledge the support of NVIDIA® Corporation with the donation of the Titan X GPU used for this research.

References

- C. D. Lehman et al., "Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study," *Radiology* **244**(2), 381–388 (2007).
- C. Kuhl, "The current status of breast MR imaging part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice," *Radiology* **244**(2), 356–378 (2007).
- C. K. Kuhl et al., "Abbreviated breast magnetic resonance imaging (MRI): first postcontrast subtracted images and maximum-intensity projection—a novel approach to breast cancer screening with MRI," *J. Clin. Oncol.* **32**(22), 2304–2310 (2014).
- V. L. Mango et al., "Abbreviated protocol for breast MRI: are multiple sequences needed for cancer detection?" *Eur. J. Radiol.* **84**(1), 65–70 (2015).
- C. D. Lehman et al., "Accuracy and interpretation time of computer-aided detection among novice and experienced breast MRI readers," *Am. J. Roentgenol.* **200**(6), W683–W689 (2013).
- S. Vreemann et al., "The performance of MRI screening in the detection of breast cancer in an intermediate and high risk screening program," in *Annual Meeting of the Int. Society for Magnetic Resonance in Medicine* (2016).
- E. B. Pages et al., "Undiagnosed breast cancer at MR imaging: analysis of causes," *Radiology* **264**(1), 40–50 (2012).
- K. Yamaguchi et al., "Breast cancer detected on an incident (second or subsequent) round of screening MRI: MRI features of false-negative cases," *Am. J. Roentgenol.* **201**(5), 1155–1163 (2013).
- D. M. Renz et al., "Detection and classification of contrast-enhancing masses by a fully automatic computer-assisted diagnosis system for breast MRI," *J. Magn. Reson. Imaging* **35**(5), 1077–1088 (2012).
- A. Vignati et al., "Performance of a fully automatic lesion detection system for breast DCE-MRI," *J. Magn. Reson. Imaging* **34**(6), 1341–1351 (2011).
- A. Gubern-Mérida et al., "Automated localization of breast cancer in DCE-MRI," *Med. Image Anal.* **20**(1), 265–274 (2015).
- Y.-C. Chang et al., "Computerized breast lesions detection using kinetic and morphologic analysis for dynamic contrast-enhanced MRI," *Magn. Reson. Imaging* **32**(5), 514–522 (2014).
- A. Gubern-Mérida et al., "Automated detection of breast cancer in false-negative screening MRI studies from women at increased risk," *Eur. J. Radiol.* **85**(2), 472–479 (2016).
- M. U. Dalmış et al., "A computer-aided diagnosis system for breast DCE-MRI at high spatiotemporal resolution," *Med. Phys.* **43**(1), 84–94 (2016).
- Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
- G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.* **42**, 60–88 (2017).
- V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA* **316**(22), 2402–2410 (2016).
- A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature* **542**(7639), 115–118 (2017).
- B. E. Bejnordi et al., "Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images," in *IEEE 14th Int. Symp. on Biomedical Imaging (ISBI)* (2017).
- M. Ghafoorian et al., "Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities," *Sci. Rep.* **7**, 5110 (2016).
- A. Mehrtash et al., "Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks," *Proc. SPIE* **10134**, 101342A (2017).
- M. U. Dalmış et al., "Using deep learning to segment breast and fibroglandular tissue in MRI volumes," *Med. Phys.* **44**(2), 533–546 (2017).
- J. H. Moltz et al., "Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans," *IEEE J. Sel. Top. Signal Process.* **3**(1), 122–134 (2009).
- B. Platel et al., "Automated characterization of breast lesions imaged with an ultrafast DCE-MR protocol," *IEEE Trans. Med. Imaging* **33**(2), 225–232 (2014).
- S. Klein et al., "Elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**(1), 196–205 (2010).
- O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
- X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.* **9**, 249–256 (2010).
- T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude," *COURSERA Neural Networks Mach. Learn.* **4**, 26–31 (2012).
- A. Srikantha, "Symmetry-based detection and diagnosis of DCIS in breast MRI," *Lect. Notes Comput. Sci.* **8142**, 255–260 (2013).
- T. Kooi and N. Karssemeijer, "Deep learning of symmetrical discrepancies for computer-aided detection of mammographic masses," *Proc. SPIE* **10134**, 101341J (2017).
- T. Kooi and N. Karssemeijer, "Classifying symmetrical differences and temporal change for the detection of malignant masses in mammography using deep neural networks," *J. Med. Imaging* **4**(4), 044501 (2017).
- M. Niemeijer et al., "On combining computer-aided detection systems," *IEEE Trans. Med. Imaging* **30**(2), 215–223 (2011).
- A. A. A. Setio et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge," *Med. Image Anal.* **42**, 1–13 (2016).
- N. Camarlinghi, "Automatic detection of lung nodules in computed tomography images: training and validation of algorithms using public research databases," *Eur. Phys. J. Plus* **128**, 1–21 (2013).
- A. A. A. Setio et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE Trans. Med. Imaging* **35**(5), 1160–1169 (2016).
- N. Camarlinghi et al., "Combination of computer-aided detection algorithms for automatic lung nodule identification," *Int. J. Comput. Assisted Radiol. Surg.* **7**(3), 455–464 (2012).
- B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Trans. Biomed. Eng.* **59**(6), 1720–1726 (2012).
- B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, CRC Press, New York (1994).
- Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).

Mehmet Ufuk Dalmış received his master's degree in biomedical engineering from Bogazici University in 2013, with his thesis titled "Similarity and Consistency Analysis of Functional Connectivity Maps." Since then, he is a PhD candidate at the Diagnostic Image Analysis Group of the RadboudUMC, Nijmegen, the Netherlands. His PhD study is focused on application of machine learning and deep learning techniques for automated analysis of breast MRI.

Suzan Vreemann is a PhD candidate at the Diagnostic Image Analysis Group within the Department of Radiology and Nuclear Medicine at the Radboud University Medical Center. She graduated in technical medicine at the University of Twente in Enschede. She started her PhD project in February 2014, focusing on the integrated classification of breast cancers in the EU-funded VPH-PRISM.

Thijs Kooi is a PhD candidate at the Department of Radiology. He received his MSc degree in artificial intelligence from the University of Amsterdam in 2012. He worked as a visiting research student at Keio University, Japan, for 8 months and as a research assistant at the A*STAR Bioinformatics Institute in Singapore for 6 months.

In April 2013, he started his PhD on computer aided diagnosis of breast cancer at DIAG.

Ritse M. Mann is a dedicated breast and interventional radiologist. He is a scientific leader of the clinical breast research line that aims at clinical validation of novel strategies for breast cancer detection and treatment. He is a member of the executive board of the European Society of Breast Imaging and its scientific and young club committees, and first author of both guidelines on breast MRI issued by this organization in 2008 and 2015, respectively.

Nico Karssemeijer is a PhD candidate at the Department of Radiology. He received his MSc degree in artificial intelligence from the

University of Amsterdam in 2012. He worked as a visiting research student at Keio University, Japan, for 8 months and as a research assistant at the A*STAR Bioinformatics Institute in Singapore for 6 months. In April 2013, he started his PhD on computer-aided diagnosis of breast cancer at DIAG.

Albert Gubern-Mérida received his joint PhD in the Diagnostic Image Analysis Group (DIAG) at Radboud University (the Netherlands) and in the University of Girona (Spain) on the topic of automated analysis of magnetic resonance imaging of the breast. His research at DIAG focuses on the development of computer-aided detection systems to improve reading of digital breast tomosynthesis images as well as other breast imaging modalities.