BREAST IMAGING
Iran J Radiol. 2022 January; 19(1):e120758. doi: 10.5812/iranjradiol-120758.
Published online 2022 April 25. **Research Article**

Microcalcification Detection in Mammograms Using Deep Learning

Mahmoud Shiri Kahnouei [1], Masoumeh Giti [2], Mohammad Ali Akhaee [3] and Ali Ameri [1,*]

[1]Department of Biomedical Engineering, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Imam Khomeini Complex Hospital, Tehran University of Medical Sciences, Tehran, Iran
[3]Department of Electrical Engineering, University of Tehran, Tehran, Iran

[*]*Corresponding author*: Department of Biomedical Engineering, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Email: aliameri86@gmail.com

Received 2021 October 30; **Revised** 2022 March 21; **Accepted** 2022 March 29.

Abstract

Background: Mammography is the most reliable and popular method in the clinical diagnosis of breast cancer. Calcifications are subtle lesions in mammograms that can be cancerous and difficult to detect for radiologists. Computer-aided detection (CAD) can help radiologists identify malignant lesions.
Objectives: This study aimed to propose a deep learning based CAD system for detecting calcifications in mammograms.
Patients and Methods: A total of 815 in-house mammograms were collected from 204 women undergoing screening mammography. Calcifications in the mammograms were annotated by specialists. Each mammogram was divided into patches of fixed size, and then, patches containing calcifications were extracted, along with the same number of normal patches. A ResNet-50 Convolutional Neural Network (CNN) was trained for classification of patches into normal and calcification groups using training data and then the performance of the trained CNN was tested with new test data.
Results: The proposed patch learning approach (PLA) showed a classification accuracy of 96.7% in the binary classification of patches. Therefore, it could detect calcification regions in a given mammogram. The PLA achieved sensitivity and specificity of 96.7% and 96.7%, respectively, with an area under the curve of 98.8%.
Conclusion: The present results highlighted the efficacy of the proposed PLA, especially for limited training data. Direct comparison with previous studies is not possible due to differences in datasets. Nevertheless, the PLA accuracy in detecting calcifications was higher than that of deep learning based CAD systems in previous studies. The effective performance of PLA may be attributed to the manual removal of uninformative patches, as they were not used in the training set.

Keywords: Breast Cancer, Mammography, CAD, Convolutional Neural Network, Deep Learning

1. Background

Breast cancer is the most common cancer and the second leading cause of death among women (1). Mammography is the primary method for examining changes in the breast tissue (2). It is recognized as the most common diagnostic method for detecting breast cancer in an early stage (3). Reduction of false positive results in cancer detection is important in therapeutic processes (4), as they can impose significant burdens on patients, including high cost, waste of time, and psychological stress (5). In developing and developed countries, the accuracy of detecting cancerous lesions is less than 50% and above 80%, respectively (6). This difference is partially due to the use of computer-aided detection (CAD) systems in developed countries.

With the advancement of medical equipment technologies, various approaches have been proposed for mammographic CAD systems (7). Since 2000, advances in diagnostic digital mammography have increased the accuracy of breast cancer detection and reduced the associated deaths (8). However, detection of suspected abnormalities is a difficult task, even for experienced radiologists. The small size of a lesion compared to the large size of a mammogram is an important dichotomy in image processing techniques for cancer detection (9). The reason for the large size of X-ray mammograms is the need for detection of very small calcification particles (10). Overall, identifying calcifications in mammograms is a subtle and time-consuming task, which causes eye strain and reduces the detection accuracy, resulting in the radiologists' error over time (11).

To reduce human errors, today, mammography CAD technologies are employed by radiologists to find suspected lesions. CAD systems can improve the rate of breast

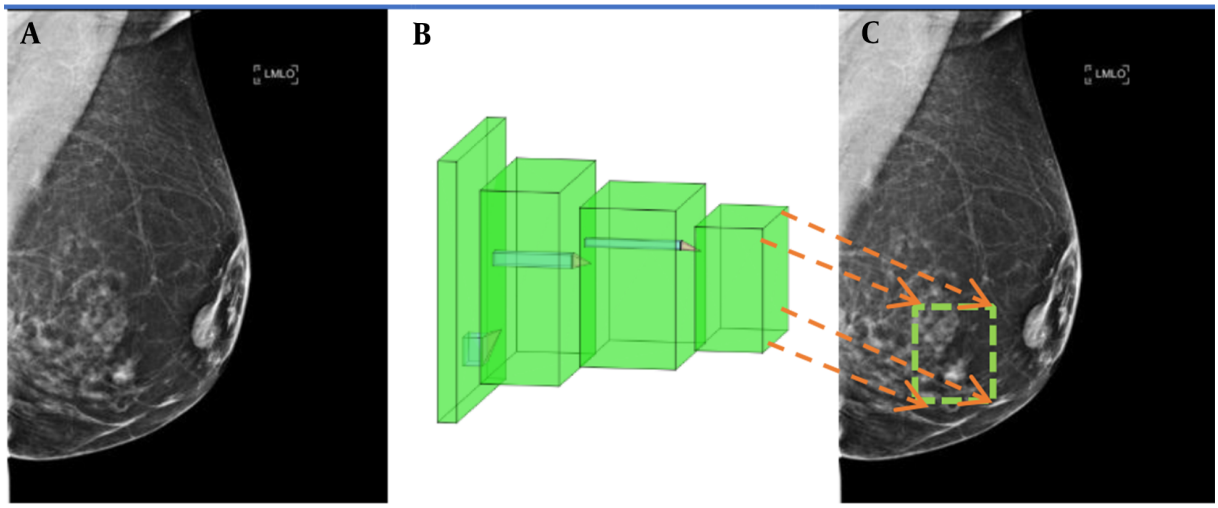

Figure 1. The conceptual workflow of a region-based convolutional neural network (CNN) for computer-aided detection (CAD). A, Input mammogram; B, CNN core for object detection; C, Output mammogram with the marked region of a lesion.

cancer diagnosis up to 20% (12). These systems often employ common deep learning-based region detection algorithms, including convolutional neural network (CNN)-based models, namely, region-based convolutional neural network (R CNN), Fast-R-CNN, Faster-R-CNN (12), and You Only Look Once (YOLO) algorithm (13). Figure 1 illustrates the conceptual workflow of region detection models, in which mammograms are fed into the network as input, and a box surrounding the lesion is marked at the output.

Common CNN structures, such as ResNet (14), VGG (15), and GoogleNet (16), which have been trained for object recognition on the ImageNet (17) database, are used in these region detection algorithms. The input size of these CNN structures is 224 $\times$ 224 pixels, whereas the size of X-ray mammograms is 2,560 $\times$ 3,328 (170 times larger than the CNN input size). Therefore, data preparation needs to be performed by downsampling to reduce the image dimensions. This resizing, however, results in the loss of informative pixels. However, empirically, only small patches in a mammogram contain lesions, and mammogram resizing blurs or eliminates them. An illustration of the negative impact of mammogram resizing on the CNN input is shown in Figure 2. To avoid this problem, this study proposed a novel alternative approach for detecting microcalcifications in mammograms.

2. Objectives

This study aimed to propose a CAD system for automated calcification detection in mammograms by training a model with an in-house dataset.

3. Patients and Methods

3.1. Dataset

A total of 815 mammograms were collected from 204 women (age: 51.43 $\pm$ 10.61 years), who were referred to Dr. Gity Imaging Center and Imam Khomeini Hospital Mammography Center during 2019 - 2020 (some participants were referred to these centers more than once). The left and right craniocaudal (CC) and mediolateral oblique (MLO) views were considered as separate images. Some women visited the centers more than once for screening, and for some women, only one breast or one view was collected. The mammography device was the Selenia Dimensions Mammography System (Hologic Inc., USA). This study was approved by the ethics board of Shahid Beheshti University of Medical Sciences, Tehran, Iran. Calcifications were annotated in the mammograms by specialists. The mammograms included CC and MLO views.

3.2. Patch Learning

In this study, each mammogram was divided into non-overlapping, fixed-size patches, which were used as the input to a CNN. By using this approach, the resolution of the

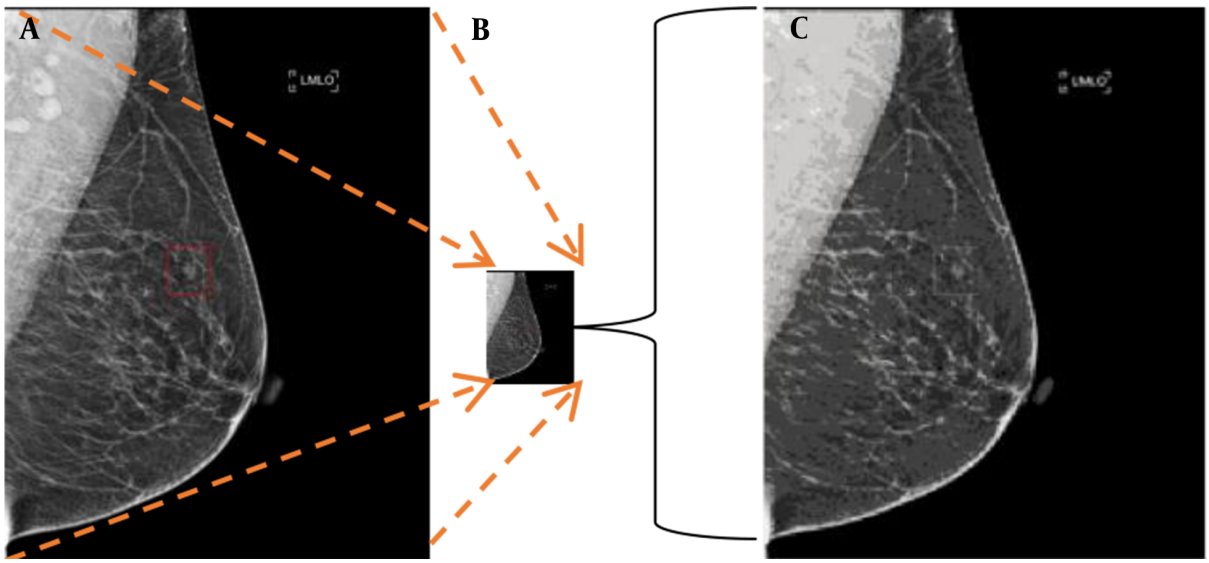

Figure 2. Mammogram resizing of convolutional neural network (CNN) input. A, Original mammogram with a size of 2560 × 3328; B, Resized mammogram with a size of 224 × 224; C, A magnified display of the resized image, showing blurred and destroyed calcification patterns.

lesion image was not deteriorated, thereby enhancing the detection accuracy. For each patch, the ground truth label was determined based on the annotations by an experienced radiologist. If a patch overlapped an annotated region, it was labelled as a calcification; otherwise, it was labelled as normal. Besides, training a CNN with patches as the input has the advantage of enlarging the training set, which improves the model training, especially considering the scarcity of mammogram datasets. We refer to the proposed approach as the patch learning approach (PLA). However, classification of malignant versus benign lesions is not within the scope of this study.

The proposed method aimed to detect and localize calcifications to help radiologists identify suspicious lesions. First, all images were divided into 224 × 224 patches. Subsequently, all patches containing calcifications were extracted from all mammograms, where 536 patches were obtained. It should be noted that if a calcification region spanned two or more adjacent patches, each of the adjacent patches was labelled as a calcification. Normal patches were extracted from all mammograms, which either contained or did not contain calcifications in other patches. Since the total number of normal patches was much higher than that of calcification patches, to achieve class size balance, a total of 536 normal patches were randomly selected from all available normal patches, and the rest of normal patches were discarded. It should be noted that an imbalance in the class size of the training set leads to classifier bias, which may degrade classification performance. Moreover, in this study, we used balanced class sizes in the test set to obtain significant performance results.

To construct the training and test sets, 100 patches, which were randomly selected from normal and calcification classes (a total of 200), were considered as the test set, while the rest of them (436 mammograms from each class) were used as the training set. The ground truth labels (based on the radiologist's annotation) were used to train the CNN with the training set. On the test data, the CNN performed a binary classification to classify patches into normal and calcification groups. The classification accuracy was determined by comparing the CNN output with the ground truth labels.

An annotated calcification in a mammogram is shown in Figure 3A, and a magnified image of the marked patch is displayed in Figure 3B. As shown in Figure 3, calcification patches often occupy less than 2% of the mammogram size; therefore, the standard procedure for resizing mammograms (reduction of image size by downsampling reduces the image resolution and results in information loss) blurs calcification spots and must be avoided.

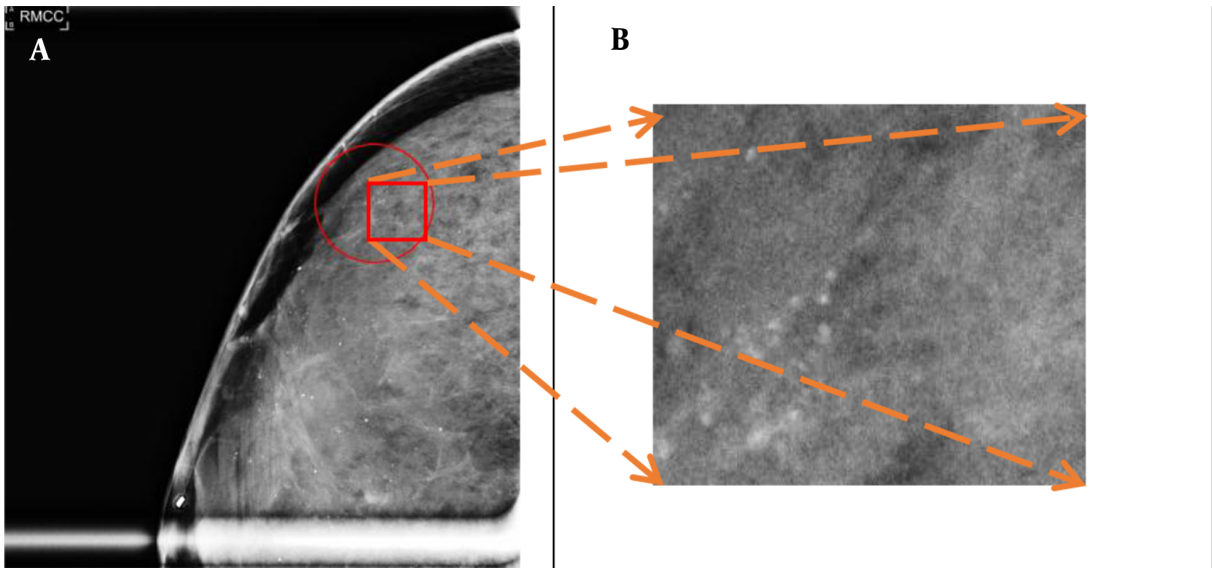

Figure 3. A, A calcification patch annotated in a mammogram; and B, A magnified lesion patch.

3.3. CNN Structure

A ResNet50 model was used for the binary classification of patches with a size of 224 × 224 pixels. This model is a modern CNN, which has been trained on millions of images and achieved excellent performance in object recognition (14). Accordingly, in this study, it was selected as our pretrained network. The ResNet50 model was fine tuned with our mammogram patches by freezing the first 100 layers. The Adam optimization method, with a 0.0001 learning rate, was selected with a binary cross entropy loss function. A batch size of 32 with 120 epochs was used for training the CNN. Computation was conducted using a TensorFlow2 platform on a computer with Nvidia' GeForce GTX 1080 GPU. The workflow of the proposed method for detecting calcifications in a new mammogram is presented in Figure 4.

4. Results

The plots of training and validation accuracy and loss during CNN training are shown in Figure 5 (5% of the training set was selected as the validation set). The close lines in the plots of training and validation indicate the CNN training efficacy. Our proposed PLA achieved an accuracy of 96.7% for the binary classification of 200 test patches. The sensitivity and specificity of PLA were 96.7% and 96.7%, respectively.

The receiver operating characteristic (ROC) curve is plotted in Figure 6, where the operating point is shown with a red dot. The area under the ROC curve (AUC) was 98.8, indicating the effective classification of the model. Generally, the operating point is selected depending on the preference, as there is a trade off between sensitivity and specificity; in other words, sensitivity can be increased at the expense of specificity, and vice versa. In this study, an operating point that maximized the classification accuracy was selected. This point was found as the intersection of the ROC curve with a line with a slope of 1 (balanced classes) and maximum y-intercept. Figure 7 indicates the performance of PLA on two representative test mammograms from the digital database for screening mammography (DDSM) dataset and our in house dataset, respectively.

By dividing the test set into three age ranges, including 31 - 45 years, 46 - 60 years, and 61 - 75 years, the model performance was calculated for each age range, as listed in Table 1. Based on the results, performance was approximately 97% and similar for all age groups.

5. Discussion

Daily reading of numerous mammograms, the majority of which are often normal, is a tedious task and may cause fatigue in radiologists, which in turn increases the risk of missing abnormalities. Therefore, there is a great

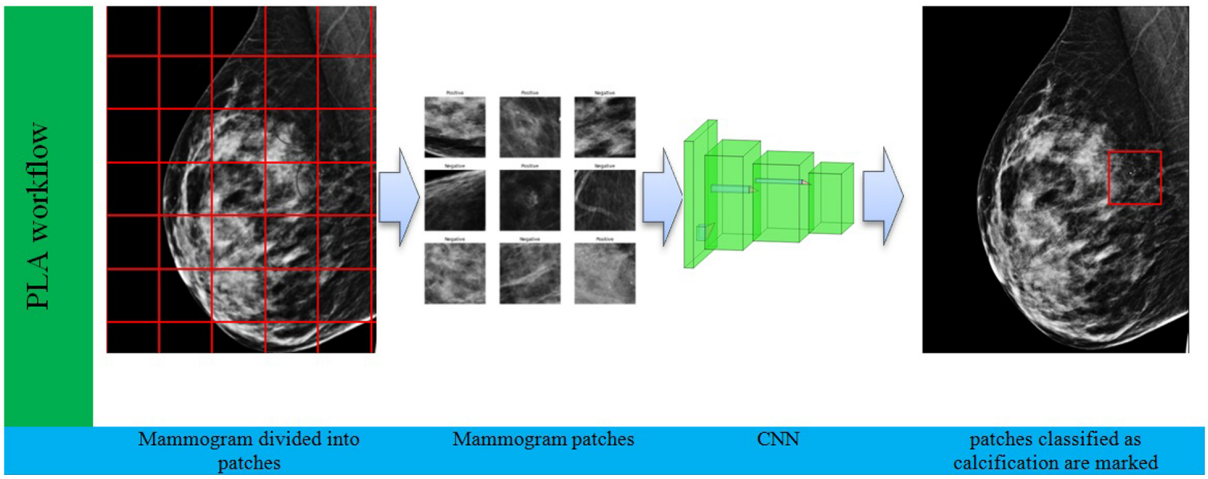

Figure 4. The mammogram is first divided into patches with a size of 224 × 224 pixels. Next, each patch is separately fed into the convolutional neural network (CNN) for classification. The CNN classifies each patch into either normal or calcification. Finally, patches classified as calcification are marked.

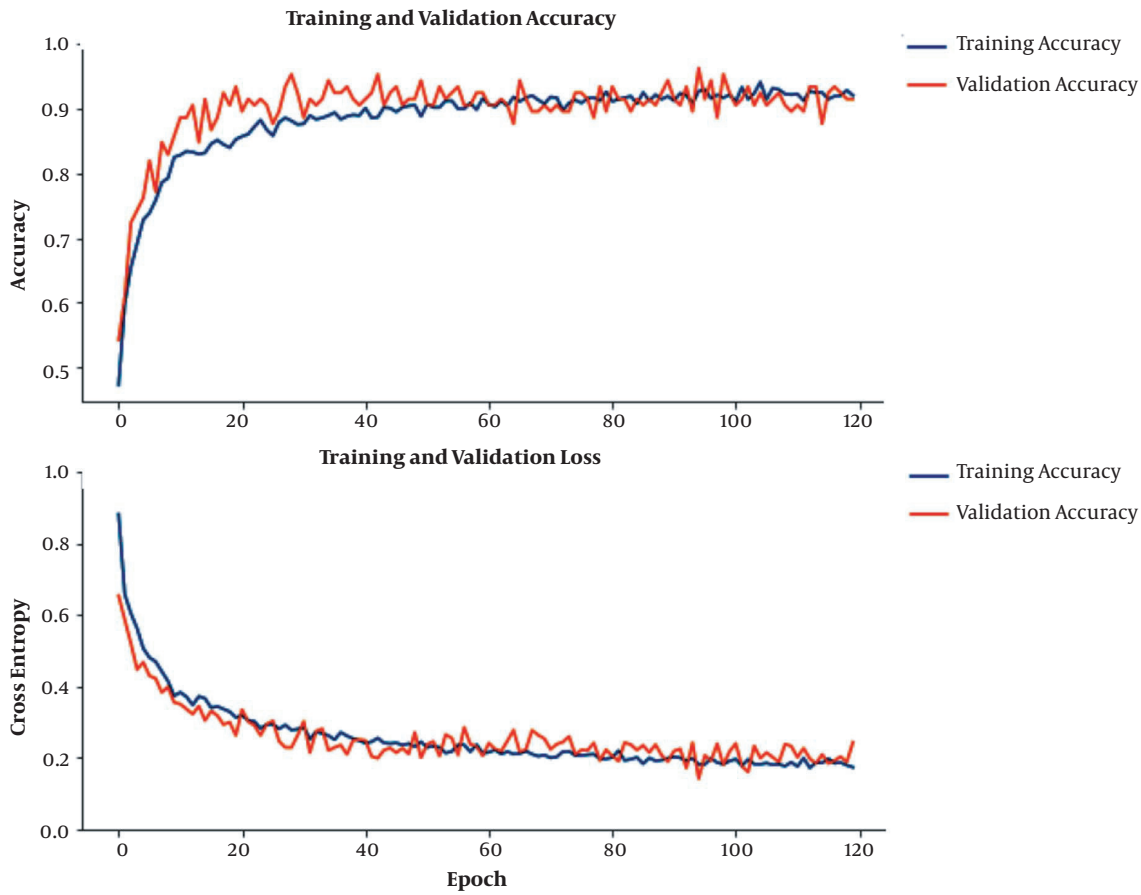

Figure 5. The accuracy and loss plots during training for the training and validation sets.

Table 1. The Classification Accuracy of the Model for Each Age Group (Along with the Number of Patches of Test Data in Each Group)

Age group (y)	Population	Classification accuracy
31 - 45	58	96.4
46 - 60	79	96.6
61 - 75	63	97.3

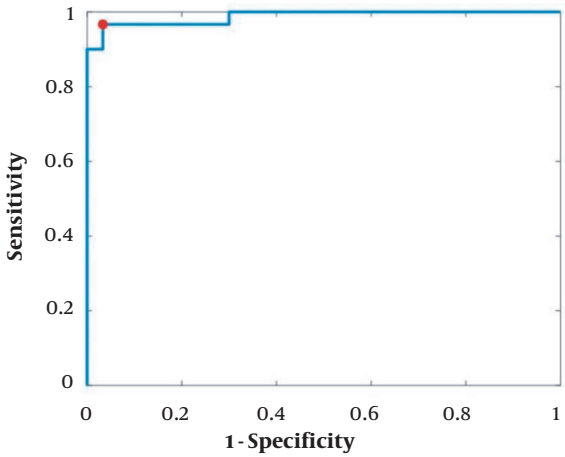

Figure 6. The receiver operating characteristic (ROC) curve for the convolutional neural network (CNN) classification of normal and calcification patches across 200 test patches. The operating point is shown with a red dot.

demand for automated detection and localization of suspicious lesions, as it can avoid missing these lesions. This study proposed a method for automated detection; however, determining the malignancy of lesions is outside the scope of this study. In our future study, we will investigate a fully automated diagnosis system for determining the malignancy of detected lesions.

Recently, deep learning has been applied in medical imaging studies, such as mammography (18). Before the introduction of deep learning, other types of machine learning (ML) methods (19) were common for detecting lesions in mammograms. These ML methods require feature engineering, which is difficult and time consuming. Feature engineering refers to the process of designing and extracting relevant and useful representations from raw data. These features need to be designed by human experts. Previous studies on ML-based detection of lesions in mammograms have used features, such as wavelet (20), curvelet (20), Fourier transform (21), and edge gradient analysis (22). These features have been also used as the input to a classifier to detect or classify lesions (23). However, because it is difficult to find perfect features, the performance of ML methods is often inferior to that of deep learning methods. In deep learning, the network automatically learns useful features so that there is no need for feature engineering. Therefore, deep learning methods can be applied directly in mammograms without any preprocessing.

The CNNs are the most commonly used deep learning frameworks in mammography CAD systems, which have shown promising performance in detecting cancerous lesions. However, large datasets are needed to train CNNs, and formation of large specialist annotated mammogram datasets is expensive and time consuming. Suspected lesions in mammograms often occupy less than 2% of image pixels. Therefore, the bulk of a mammogram does not contain useful information for training a deep learning model. In previous studies, such as a study by Agarwal et al. (24), the whole image was fed into CNN models, which increased the training time substantially, while most of the data was not informative. As a solution, the proposed PLA divides the image into fixed-size patches, and only suspected patches, along with the same number of normal patches, were fed into the CNN for training. This strategy substantially reduced the training time, as only informative patches, which comprise a small percentage of each mammogram, were used for training the CNN.

The proposed PLA also has the advantage of being adaptable to various sizes of mammograms, as it operates on fixed-size patches of images. This allows the PLA to be trained with one dataset and tested by another with a different mammogram size. Table 2 lists the performance (AUC) of deep learning-based CAD systems for detecting calcifications in previous studies compared to our proposed PLA; the models and datasets used in these studies are also demonstrated. Our method outperformed these previous approaches; however, it should be noted that a direct comparison is not possible due to differences in datasets. The higher performance of our system may be attributed to the patch learning algorithm.

In conclusion, the results of this study highlighted the efficacy of our PLA. Future studies are suggested to focus on the application of this approach for detecting both masses and calcifications in mammograms.

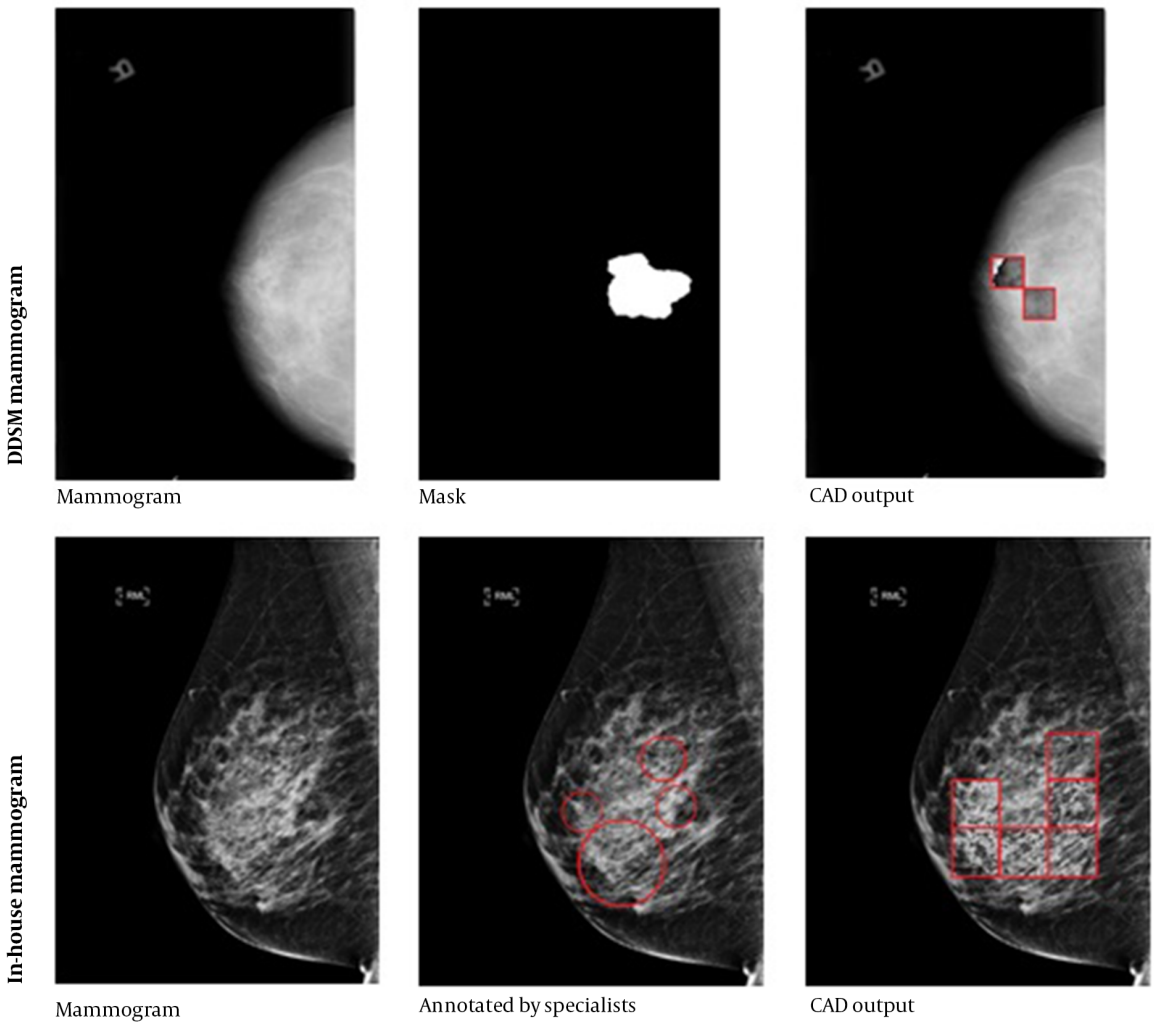

Figure 7. The performance of the patch learning approach (PLA) on two representative test mammograms from the digital database for screening mammography (DDSM) dataset (first row) and our in house dataset (second row). The middle column shows the annotations by specialists, and the third column shows the marked patches by the proposed computer-aided detection (CAD) system.

Table 2. Comparison of Deep Learning-Based Mammography Studies

Studies	Databases	AUC
Agnes et al. (25)	Mini-MIAS	96%
Shen et al. (26)	CBIS-DDSM	91%
Wang et al. (27)	DDSM, INbreast , and MIAS	88%
Herent et al. (28)	MIAS, INbreast, and Journees Francophones de Radiologie	81%
Our proposed PLA	In-house dataset and CBIS-DDSM	98%

Abbreviations: AUC, area under the ROC curve; Mini-MIAS, Mini-Mammographic Image Analysis Society; CBIS-DDSM, curated breast imaging subset of the digital database for screening mammography; INbreast, full-field digital mammographic database; MIAS, Mammographic Image Analysis Society.

Footnotes

Authors' Contribution: Study concept and design, analysis and interpretation of data, and deep learning structure design, A.A.; Analysis and interpretation of data and critical revision of the manuscript for important intellectual content, M.G.; Analysis and interpretation of data, deep learning structure design, AI conception, statistical analysis, and drafting of the manuscript, M.Sh.; and Statistical analysis, M.A.A.

Conflict of Interests: Masoumeh Giti is a faculty member of Tehran University of Medical Sciences and also an editorial board member of the Iranian Journal of Radiology.

Data Reproducibility: The dataset presented in the study is available on request from the corresponding author during submission or after its publication. The data are not publicly available due to legal constraints.

Ethical Approval: This study was approved under the ethical approval code of IR.SBMU.MSP.REC.1398.840 (link to the ethical approval code: ethics.research.ac.ir/ProposalCertificateEn.php?id = 102536).

Funding/Support: This study was partly supported by a grant from Shahid Beheshti University of Medical Sciences (19049).

Informed Consent: The "Inform Consent" form is uploaded in the supplementary files during the submission. (file name: Informed Consent 120758.pdf).

References

1. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology*. 2009;**253**(3):641–51. doi: 10.1148/radiol.2533082308. [PubMed: 19864507]. [PubMed Central: PMC2786197].
2. Azam S, Eriksson M, Sjolander A, Gabrielson M, Hellgren R, Czene K, et al. Mammographic microcalcifications and risk of breast cancer. *Br J Cancer*. 2021;**125**(5):759–65. doi: 10.1038/s41416-021-01459-x. [PubMed: 34127810]. [PubMed Central: PMC8405644].
3. Halladay JR, Yankaskas BC, Bowling JM, Alexander C. Positive predictive value of mammography: comparison of interpretations of screening and diagnostic images by the same radiologist and by different radiologists. *AJR Am J Roentgenol*. 2010;**195**(3):782–5. doi: 10.2214/AJR.09.2955. [PubMed: 20729460]. [PubMed Central: PMC4451561].
4. Daniaux M, Gruber L, Santner W, Czech T, Knapp R. Interval breast cancer: Analysis of occurrence, subtypes and implications for breast cancer screening in a model region. *Eur J Radiol*. 2021;**143**:109905. doi: 10.1016/j.ejrad.2021.109905. [PubMed: 34403883].
5. Mendelson EB. Artificial Intelligence in Breast Imaging: Potentials and Limitations. *AJR Am J Roentgenol*. 2019;**212**(2):293–9. doi: 10.2214/AJR.18.20532. [PubMed: 30422715].
6. Trieu PDY, Lewis SJ, Li T, Ho K, Wong DJ, Tran OTM, et al. Improving radiologist's ability in identifying particular abnormal lesions on mammograms through training test set with immediate feedback. *Sci Rep*. 2021;**11**(1):9899. doi: 10.1038/s41598-021-89214-3. [PubMed: 33972611]. [PubMed Central: PMC8110801].
7. Bhushan A, Gonsalves A, Menon JU. Current State of Breast Cancer Diagnosis, Treatment, and Theranostics. *Pharmaceutics*. 2021;**13**(5). doi: 10.3390/pharmaceutics13050723. [PubMed: 34069059]. [PubMed Central: PMC8156889].
8. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Semin Cancer Biol*. 2021;**72**:214–25. doi: 10.1016/j.semcancer.2020.06.002. [PubMed: 32531273].
9. Niu S, Huang J, Li J, Liu X, Wang D, Wang Y, et al. Differential diagnosis between small breast phyllodes tumors and fibroadenomas using artificial intelligence and ultrasound data. *Quant Imaging Med Surg*. 2021;**11**(5):2052–61. doi: 10.21037/qims-20-919. [PubMed: 33936986]. [PubMed Central: PMC8047381].
10. Dandu RV. Storage media for computers in radiology. *Indian J Radiol Imaging*. 2008;**18**(4):287–9. doi: 10.4103/0971-3026.43838. [PubMed: 19774182]. [PubMed Central: PMC2747448].
11. Arancibia Hernández PL, Taub Estrada T, López Pizarro A, Díaz Cisternas ML, Sáez Tapia C. Calcificaciones mamarias: descripción y clasificación según la 5.a edición BI-RADS. *Rev Chil Radiol*. 2016;**22**(2):80–91. doi: 10.1016/j.rchira.2016.06.004.
12. Bargallo X, Santamaria G, Del Amo M, Arguis P, Rios J, Grau J, et al. Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program. *Eur J Radiol*. 2014;**83**(11):2019–23. doi: 10.1016/j.ejrad.2014.08.010. [PubMed: 25193778].
13. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv*. 2017;**Preprint**.
14. Targ S, Almeida D, Lyman K. Resnet in resnet: Generalizing residual architectures. *arXiv*. 2016;**Preprint**.
15. Sengupta A, Ye Y, Wang R, Liu C, Roy K. Going Deeper in Spiking Neural Networks: VGG and Residual Architectures. *Front Neurosci*. 2019;**13**:95. doi: 10.3389/fnins.2019.00095. [PubMed: 30899212]. [PubMed Central: PMC6416793].
16. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the Inception Architecture for Computer Vision. *Conference on Computer Vision and Pattern Recognition*. 2012; Tennessee, USA. IEEE; 2012. p. 2818–26.
17. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;**25**:1097–105.
18. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R. Deep learning in mammography and breast histology, an overview and future trends. *Med Image Anal*. 2018;**47**:45–67. doi: 10.1016/j.media.2018.03.006. [PubMed: 29679847].
19. te Brake GM, Karssemeijer N, Hendriks JH. An automatic method to discriminate malignant masses from normal tissue in digital mammograms. *Phys Med Biol*. 2000;**45**(10):2843–57. doi: 10.1088/0031-9155/45/10/308. [PubMed: 11049175].
20. Ghazali KH, Mansor MF, Mustafa MM, Hussain A. Feature Extraction Technique using Discrete Wavelet Transform for Image Classification. *5th Student Conference on Research and Development*. Selangor, Malaysia. IEEE; 2007. p. 1–4.

21. Zhang Y, Chen S, Wang S, Yang J, Phillips P. Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine. *Int J Imaging Syst Technol*. 2015;**25**(4):317–27. doi: 10.1002/ima.22144.
22. Feichtenhofer C, Fassold H, Schallauer P. A Perceptual Image Sharpness Metric Based on Local Edge Gradient Analysis. *IEEE Signal Processing Letters*. 2013;**20**(4):379–82. doi: 10.1109/lsp.2013.2248711.
23. Kozegar E, Soryani M, Minaei B, Domingues I. Assessment of a novel mass detection algorithm in mammograms. *J Cancer Res Ther*. 2013;**9**(4):592–600. doi: 10.4103/0973-1482.126453. [PubMed: 24518702].
24. Agarwal R, Diaz O, Yap MH, Llado X, Marti R. Deep learning for mass detection in Full Field Digital Mammograms. *Comput Biol Med*. 2020;**121**:103774. doi: 10.1016/j.compbiomed.2020.103774. [PubMed: 32339095].
25. Agnes SA, Anitha J, Pandian SIA, Peter JD. Classification of Mammogram Images Using Multiscale all Convolutional Neural Network (MA-CNN). *J Med Syst*. 2019;**44**(1):30. doi: 10.1007/s10916-019-1494-z. [PubMed: 31838610].
26. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Sci Rep*. 2019;**9**(1):12495. doi: 10.1038/s41598-019-48995-4. [PubMed: 31467326]. [PubMed Central: PMC6715802].
27. Wang X, Liang G, Zhang Y, Blanton H, Bessinger Z, Jacobs N. Inconsistent Performance of Deep Learning Models on Mammogram Classification. *J Am Coll Radiol*. 2020;**17**(6):796–803. doi: 10.1016/j.jacr.2020.01.006. [PubMed: 32068005].
28. Herent P, Schmauch B, Jehanno P, Dehaene O, Saillard C, Balleyguier C, et al. Detection and characterization of MRI breast lesions using deep learning. *Diagn Interv Imaging*. 2019;**100**(4):219–25. doi: 10.1016/j.diii.2019.02.008. [PubMed: 30926444].