Iranian Journal of Radiology
Volume 19 · Issue 1 · e120758

OPEN PEER REVIEW

Microcalcification Detection in Mammograms Using Deep Learning

Mahmoud Shiri Kahnouei, Masoumeh Gity, Mohammad Ali Akhaee, Ali Ameri

SUBMITTED	REVISED	ACCEPTED	DOI
30 Oct 2021	21 Mar 2022	29 Mar 2022	10.5812/iranjradiol-120758

About this report

This open peer review (OPR) report shows the public comments made on this article during peer review, by the reviewers, the associate editor and the editor-in-chief, for every revision round, together with the authors' replies. Comments appear only with the consent of the people who wrote them. Reviewers who chose to sign their report are named; the others are shown as "Referee". Comments are edited for language only, and confidential remarks to the editors are not included.

More about open peer review: brieflands.com/knowledgebase/opr

Contents

1 **Original submission** Referee 1, Referee 2, Associate Editor, Editor-in-Chief 2
2 **Revision 1** Editor-in-Chief 4
Authors' reply to reviewers 5
Reviewer and editor comments 11
3 **Revision 2** 12
Authors' reply to reviewers 13
Reviewer and editor comments 16
4 **Revision 3** Editor-in-Chief 17
Authors' reply to reviewers 18
Reviewer and editor comments 21
5 **Revision 4** Editor-in-Chief 22
Authors' reply to reviewers 23
Reviewer and editor comments 25

1 Original submission

4 comments · 11 Nov 2021 – 27 Dec 2021

Referee 1 Reviewer

MINOR REVISION 11 Nov 2021

Dear authors
Thank you for submitting your article to IJR.
I have some questions and comments.
1-Please do not use abbreviations in the title.
2-Please add the features of the mammography device.
3-How many radiologists interpreted the images?
4-You had 214 patients and 815 mammograms.
It seems that the left and right cc and MLO images were each considered a separate image.
Please add this information to the study methods.
5-What were the results of the biopsies? How many calcifications were malignant? Can you say how effective these methods are for benign and malignant calcifications?
Best regards

Referee 2 Reviewer

MINOR REVISION 23 Dec 2021

The article is relevant and original.
Good novelty.

Does this patch learning approach (PLA) determine whether the microcalcifications are benign or malignant? Or does it only detect a calcification patch?
Please explain further.

Associate Editor

MAJOR REVISION 27 Dec 2021

1-Please add the full terms for abbreviations at first use [separately in the abstract and the paper] [for example, AUC].

2-Some parts of the background are, in fact, content related to the method section. They should be moved to the method section.

3-How did the authors select the normal patches from all patches? By randomization? Please explain.

4-Did the authors select the normal patches from mammograms without calcification? Or did they select them from non-calcification regions of mammograms with calcification? Or both of these scenarios? Please explain.

5-Is it possible for a region of microcalcification to be located in two adjacent patches?

6-What is the “standard procedure of resizing mammograms”?

7-What is the gold standard? If the radiologist's determination is the gold standard, why did the authors mention that some patients were confirmed by pathology? What was confirmed by pathology? Calcification, cancer, or the lesion diagnosis?

8-Detection of calcification seems easy. In fact, what is challenging in breast imaging is categorization of the type of microcalcification [not the presence of microcalcification]. Is it so important to design a CAD for detection of microcalcification?

9-Is ResNet50 a valid program? Is there any paper in this regard?

10-What is the meaning of binary classification? If this means classifying the patches into positive and negative patches [regarding the presence of microcalcification], it seems this should be done based on the radiologist's opinion [when we want to determine the exact nature of a patch based on the gold standard].

11-How many radiologists assessed the images? What is their reliability [inter- and intra-rater reliability]?

12-Table 2 is missing in the paper.

13-Why did the authors not select all the patches and image data?

14-What was the method of selecting or assigning the patches for the training and test sets? By randomization?

15-The discussion is so brief and does not have any references. Please complete the discussion and compare it with other similar works. Please compare the results with those of other image processing methods.

Editor-in-Chief MAJOR REVISION 27 Dec 2021

Dear authors,
Your article needs Major Revision.
Regards

2

Revision 1

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Manuscript ID: 120758 IJR
Response to Associate Editors and Reviewers

Dear Editor-in-chief,
We would like to thank the Associate Editors and reviewers for their constructive comments. We have revised the manuscript to address these comments as described in the following point-by-point responses. All changes made in the manuscript is highlighted in the revised version.

Reviewers' Comments to the Authors:

Reviewer 1:
Dear authors; Thank you for submitting your article to IJR, I have some questions and comments

Q1. Please do not use abbreviation in the title
R1. The title has been changed to: "Micro-calcification detection in mammograms using deep learning"

Q2. Please add the features of the mammography device
R2. The mammography system was Selenia Dimensions, Hologic Inc.
This information has been added in the revised manuscript.

Q3. How many radiologists have interpreted the images?
R3. An experienced radiologist read and annotated the mammograms.
This information has been added in the revised manuscript.

Q4. You had 214 patients and 815 mammograms, and it seems that the left and right cc and MLO images are considered as a separate image. Please add this in the study method.
R4. Yes, the left and right CC and MLO were considered as separate images. Some women referred more than once for screening, and for some women only one breast or one view was collected.
This information has been added in the revised manuscript.

Q5. What were the results of the biopsies? How many calcifications were malignant? can you say how effective are these methods in benign and malignant calcifications?
R5. To avoid confusions, the biopsy sentence has been deleted.
We did not collect the biopsy results, as a classification of malignant vs. benign lesions is outside the scope of this work. The goal of the proposed method is detection and localization of calcifications, to help radiologists identify suspicious lesions. In our future work, we will predict the malignancy of the detected lesions.
This information has been added in the revised manuscript.

Reviewer 2:

the article is relevant and original, good novelty

Q1. Does this patch learning approach (PLA) determine whether the microcalcifications are benign or malignant? Or only detect calcification patch, Please explain more

R1. To avoid confusions, the biopsy sentence has been deleted.

We did not collect the biopsy results, as a classification of malignant vs. benign lesions is outside the scope of this work. The goal of the proposed method is detection and localization of calcifications, to help radiologists identify suspicious lesions. In our future work, we will predict the malignancy of the detected lesions.

This information has been added in the revised manuscript.

Associate Editor's Comments:

Associate Editor 1:

Dear authors, regarding reviewers' comments the article needs minor revision.

Reviewer 1:

the article is relevant and original, good novelty

Q1. Does this patch learning approach (PLA) determine whether the microcalcifications are benign or malignant? Or only detect calcification patch, Please explain more

R1. To avoid confusions, the biopsy sentence has been deleted.

We did not collect the biopsy results, as a classification of malignant vs. benign lesions is outside the scope of this work. The goal of the proposed method is detection and localization of calcifications, to help radiologists identify suspicious lesions. In our future work, we will predict the malignancy of the detected lesions.

This information has been added in the revised manuscript

Reviewer 2:

Dear authors; Thank you for submitting your article to IJR, I have some questions and comments

Q1. Please do not use abbreviation in the title

R1. The title has been changed to: “Micro-calcification detection in mammograms using deep learning”

Q2. Please add the features of the mammography device

R2. The mammography system was Selenia Dimensions, Hologic Inc.

This information has been added in the revised manuscript.

Q3. How many radiologists have interpreted the images?

R3. An experienced radiologist read and annotated the mammograms.

This information has been added in the revised manuscript.

Q4. You had 214 patients and 815 mammograms, and it seems that the left and right cc and MLO images are considered as a separate image. Please add this in the study method.

R4. Yes, the left and right CC and MLO were considered as separate images. Some women referred more than once for screening, and for some women only one breast or one view was collected.

This information has been added in the revised manuscript.

Q5. What were the results of the biopsies? How many calcifications were malignant? can you say how effective are these methods in benign and malignant calcifications?

R5. To avoid confusions, the biopsy sentence has been deleted.

We did not collect the biopsy results, as a classification of malignant vs. benign lesions is outside the scope of this work. The goal of the proposed method is detection and localization of calcifications, to help radiologists identify suspicious lesions. In our future work, we will predict the malignancy of the detected lesions.

This information has been added in the revised manuscript.

Associate Editor 2:

Q1. Please add the full term of abbreviations for the first usage of that term [separately in abstract and paper] [for example AUC]

R1. The full terms have been added in the revised manuscript.

Q2. Some parts of background are in fact the contents related to method section. They are better to be moved to method section.

R2. A part of the Background has been moved to the Methods section in the revised manuscript.

Q3. How did the authors select the normal patches among all patches? By randomization? Please explain.

R3. For each patch, the ground truth label was determined based on the radiologist annotation. If a patch did not overlap an annotated region, then it was labelled normal.

Because normal patches were much more than calcification patches, to attain class size balance, 536 normal patches were selected randomly from all normal patches available, and the rest were discarded.

This information has been added in the revised manuscript.

Q4. did the authors select the normal patches among the mammograms without calcification? Or they selected them from non-calcification regions of mammograms with calcification? Or both of these scenarios? Please explain.

R4. The normal patches were non-calcification regions from any mammogram which either had or did not have calcifications in other patches.
This information has been added in the revised manuscript

Q5. Is it possible a region of microcalcification locate in two adjacent patches?
R5. Yes, if a calcification region spans two or more adjacent patches, each of these adjacent patches are labelled as calcification (positive).
This information has been added in the revised manuscript.

Q6. what is "standard procedure of resizing mammograms"
R6-This procedure is reducing the size of the image by downsampling which lowers the image resolution and hence results in information loss.
This information has been added in the revised manuscript.

Q7. What is the gold standard? If the radiologist determination is the gold standard, why did they mention some patients were confirmed by pathology? What was confirmed by pathology? Calcification or cancer or lesion diagnosis?
R7. To avoid confusions, the biopsy sentence has been deleted.
We did not collect the biopsy results, as a classification of malignant vs. benign lesions is outside the scope of this work. The goal of the proposed method is detection and localization of calcifications, to help radiologists identify suspicious lesions. In our future work, we will predict the malignancy of the detected lesions.
This information has been added in the revised manuscript.

Q8. Detection of calcification seems easy. In fact, what is challenging in the breast imaging is categorization of type of microcalcification [not presence of mcrocalcification] Is it so important to design a CAD for detection of microcalcification?
R8. Daily reading of numerous mammograms, where a large majority of them are often normal, is a tedious task, and tends to result in radiologist fatigue, which in turn increases the chance of missing detection of abnormalities. Therefore, there is a great demand for automated detection and localization of suspicious lesions, as it can avoid missing these lesions. This study proposed a method for automated detection, but determining the malignancy of the detected lesions were outside the scope of this work. Our future work will investigate a fully-automated diagnosis to determine the malignancy of the detected lesions.

Q9. is ResNet50 a valid program? Is there any paper in this regard?
R9-ResNet50 is a modern CNN which has achieved excellent performance in object recognition [1,2], and therefore was selected as our pretrained network in this study.

Q10. what is the meaning of binary classification? If this means classifying the patches in to positive and negative patches [regarding presence of

microcalcification], it seems this should be done based on radiologist opinion [when we want to have the exact nature of patch based on gold standard]

R10. Yes, on the test data, the CNN performed a binary classification to classify the patches into positive and negative patches. The classification accuracy was determined by comparing the CNN output with the gold standard (ground truth). The ground truth labels were based on the radiologist annotation; a patch overlapping an annotated region was labelled positive, and otherwise was labelled negative. The ground truth labels in the training set has been used to train the CNN.

Q11. How many radiologists did assess the images? What is their reliability [inter- and Intra-rater reliability]

R11. An experienced radiologist annotated the mammograms.

Q12. table 2 is missing in the paper.

R12-This is a typo; “2” has been changed to “1” in the revised version.

Q13. why did not the authors select the whole patches and data of the images?

R13. An imbalance of class sizes in the training set leads to classifier bias which may degrade the classification performance. Since normal patches were much more than calcification patches, some normal patches in the training set were discarded to attain class size balance.

We also used balanced class sizes in the test set to obtain more meaningful performance results.

Q14. what was the method of selecting or assigning the patches for training and test sets? By randomization?

R14. Yes, as stated in the following sentence of the manuscript:

Finally, 100 patches randomly selected from each class (normal or calcification), i.e. 200 in total, were chosen as test set, and the rest (436 mammograms from each class) were used as the training set.

Q15. discussion is so brief and does not have any reference. Please complete the discussion and compare with other similar works. please compare the results with other image processing methods.

R15. The following paragraphs have been added to the Discussion. Also, Table 1 has been moved to the Discussion.

Daily reading of numerous mammograms, where a large majority of them are often normal, is a tedious task, and tends to result in radiologist fatigue, which in turn increases the chance of missing detection of abnormalities. Therefore, there is a great demand for automated detection and localization of suspicious lesions, as it can avoid missing these lesions. This study proposed a method for automated detection, but determining the malignancy of the detected lesions were outside the scope of this

work. Our future work will investigate a fully-automated diagnosis to determine the malignancy of the detected lesions.

Deep learning has been very recently applied to medical images such as mammograms. However, before that, other types of machine learning (ML) methods 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 the raw data. These features have to be designed by a human expert. Previous studies on ML-based detection of lesions in mammograms used features such as wavelet, curvelet, Fourier transform, and edge-gradient analysis. These features were then used as inputs to a classifier to detect or classify lesions. 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 to mammogram images without any preprocessing.

Table 1 lists the performance (AUC) of previous work deep learning-based CAD systems for detecting calcifications, in comparison to our proposed PLA. The model and datasets used in these studies are also shown. Our method outperformed these previous approaches, but it should be noted that a direct comparison is not possible due to differences in datasets. This higher performance may be attributed to the patch learning algorithm.

2 Revision 1

1 comment · 12 Feb 2022

Editor-in-Chief

MINOR REVISION 12 Feb 2022

Dear authors,
Your article needs Minor Revision.
Regards

3

Revision 2

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Manuscript ID: 120758 IJR
Response to Associate Editors and Reviewers

Dear Editor-in-chief,
We would like to thank the Associate Editors and reviewers for their constructive comments. We have revised the manuscript to address these comments as described in the following point-by-point responses. All changes made in the manuscript is highlighted in the revised version.

Reviewers' Comments to the Authors:

Reviewer 1:
Considering the corrections made, the article is accepted.
Thank you

Reviewer 2:
Dear authors
Thank you for revision
I think this article is suitable for publication
Best regards
Thank you

Reviewer 3:
Thanks from authors for their submission to the IJR.
Totally, it can be said that this manuscript is appropriate according to scientific writing, statistical analysis, and results presentation; but it needs minor revisions as follows:
Thank you

Q1. Ethical consideration section must be written with more derailed explanation
R1. Ethical validation for the study was obtained from the SBMU ethics board (IR.SBMU.MSP.REC.1398.840). The patients' names were removed from the mammogram data to ensure confidentiality and anonymity of participants.
This information has been added in the revised manuscript.

Q2. Statistical analysis section must be added to the end of method part
R2. Because the results (outcomes for each patch: success or failure) are binary, it would not be correct to conduct statistical analyses such as ANOVA on the results

Q3. For calculation of sensitivity and specificity, a gold standard method is needed. Therefore, this issue must be added in the method part.
R3. The following information has been added in the manuscript.

In general, the choice of the operating point depends on the preference, as there is a trade-off between sensitivity and specificity, i.e. sensitivity can be increased at the expense of specificity and vice versa. In this study, we chose an operating point that maximized the classification accuracy. This point was found as the intersection of the ROC curve and a line with a slope of 1 (because the classes were balanced) which had maximum Y-intercept.

Q4. Although in practice the terms used for the horizontal and vertical axes of Figure 6 are correct; but, formally it should be used of 1-especificity for the horizontal axis and sensitivity for the vertical axis.

R4. This has been modified in the revised manuscript

Q5. The title of figures must be written below of them.

R5. This has been modified in the revised manuscript

Q6. It is better to evaluate the validity of model based on all important variables such as age.

R6. The following information has been added in the revised manuscript.

By dividing the test set into three age ranges including 31-45 years, 46-60 years, 61-75 years, the model performance for each age range was calculated as listed in Table1. Based on the results, the performance for all age groups was similar and approximately 97%.

Table 1: Classification accuracies for each age group; the number of patches of the test data in each group are also shown.

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

Reviewer 4:

Dear author

Q1. Please clarify the financial resource and mention the number grant.

Regarding the importance of “conflict of interest” clarification, corresponding author should fill the “ICMJ form” form and upload in the submission process. (The form exists in “conflict of interest” section of the site: https://brief.land/iranjradiol/knowledgebase/category/tree.html#conflicts_of_interest.html).

R1. The file will be uploaded in the submission process

Associate Editor's Comments:

Associate Editor 1:

Based on reviewers' comments the article can be accepted after minor revision.

EIC Decision:

Dear authors,

Your article needs Minor Revision.

Regards

3 Revision 2

0 comments

No public comments in this round.

4

Revision 3

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Manuscript ID: 120758 IJR
Response to Associate Editors and Reviewers

Dear Editor-in-chief,
We would like to thank the Associate Editors and reviewers for their constructive comments. We have revised the manuscript to address these comments as described in the following point-by-point responses. All changes made in the manuscript is highlighted in the revised version.

Reviewers' Comments to the Authors:

Reviewer 1:
Considering the corrections made, the article is accepted.
Thank you

Reviewer 2:
Dear authors
Thank you for revision
I think this article is suitable for publication
Best regards
Thank you

Reviewer 3:
Thanks from authors for their submission to the IJR.
Totally, it can be said that this manuscript is appropriate according to scientific writing, statistical analysis, and results presentation; but it needs minor revisions as follows:
Thank you

Q1. Ethical consideration section must be written with more derailed explanation
R1. Ethical validation for the study was obtained from the SBMU ethics board (IR.SBMU.MSP.REC.1398.840). The patients' names were removed from the mammogram data to ensure confidentiality and anonymity of participants.
This information has been added in the revised manuscript.

Q2. Statistical analysis section must be added to the end of method part
R2. Because the results (outcomes for each patch: success or failure) are binary, it would not be correct to conduct statistical analyses such as ANOVA on the results

Q3. For calculation of sensitivity and specificity, a gold standard method is needed. Therefore, this issue must be added in the method part.
R3. The following information has been added in the manuscript.

In general, the choice of the operating point depends on the preference, as there is a trade-off between sensitivity and specificity, i.e. sensitivity can be increased at the expense of specificity and vice versa. In this study, we chose an operating point that maximized the classification accuracy. This point was found as the intersection of the ROC curve and a line with a slope of 1 (because the classes were balanced) which had maximum Y-intercept.

Q4. Although in practice the terms used for the horizontal and vertical axes of Figure 6 are correct; but, formally it should be used of 1-especificity for the horizontal axis and sensitivity for the vertical axis.

R4. This has been modified in the revised manuscript

Q5. The title of figures must be written below of them.

R5. This has been modified in the revised manuscript

Q6. It is better to evaluate the validity of model based on all important variables such as age.

R6. The following information has been added in the revised manuscript.

By dividing the test set into three age ranges including 31-45 years, 46-60 years, 61-75 years, the model performance for each age range was calculated as listed in Table1. Based on the results, the performance for all age groups was similar and approximately 97%.

Table 1: Classification accuracies for each age group; the number of patches of the test data in each group are also shown.

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

Reviewer 4:

Dear author

Q1. Please clarify the financial resource and mention the number grant.

Regarding the importance of “conflict of interest” clarification, corresponding author should fill the “ICMJ form” form and upload in the submission process. (The form exists in “conflict of interest” section of the site: https://brief.land/iranjradiol/knowledgebase/category/tree.html#conflicts_of_interest.html).

R1. The file will be uploaded in the submission process

Associate Editor's Comments:

Associate Editor 1:
Based on reviewers' comments the article can be accepted after minor revision.

EIC Decision:

Dear authors,
Your article needs Minor Revision.
Regards

2:

Dear Editor-in-chief,
Thank you for your comments.

EIC Decision:

Dear authors,
Before starting the review processing of your article, it is necessary to clarify the following items:

Q1. Please clarify the financial resource and mention the number grant with the link address or upload official approval grant number.
R1. The financial resource has been added in the submission process as follows:
This study was supported by grant 19049, IR.SBMU.MSP.REC.1398.840
Link: https://ethics.research.ac.ir/ProposalCertificateEn.php?id=102536&Print=true&NoPrintHeader=true&NoPrintFooter=true&NoPrintPageBorder=true&LetterPrint=true

Q2. As some authors are faculty members of Tehran university of medical sciences or editorial board members of the IJR, this should be mentioned as conflict of interest. (for example in your paper; Masoumeh Giti is faculty member of Tehran university of medical sciences and also editorial board member of the IJR).
R2. The conflict of interest has been added in the submission process as follows:
Masoumeh Giti is a faculty member of Tehran university of medical sciences and also an editorial board member of the IJR.

Regards

4 Revision 3

1 comment · 16 Mar 2022

Editor-in-Chief

MINOR REVISION 16 Mar 2022

Dear authors,
Your article needs Minor Revision.
Regards

5

Revision 4

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Manuscript ID: 120758 IJR
Response to the Associate Editors and Reviewers

Dear Editor-in-chief,
We would like to thank the Associate Editors and reviewers for their constructive comments. We have revised the manuscript to address these comments as described in the following point-by-point responses. All changes made in the manuscript are highlighted in the revised version.

Reviewers' Comments:

Reviewer 1:

Q1. References:
References 25 to 28 do not cite in the text of the article. Please citation them.
R1. References 25 to 28 have been cited in Table 2(comparison related to mammography studies).

Associate Editor's Comments:

Associate Editor 1:
The article is accepted for publication in IJR.
Thank you.

Associate Editor 2:
120758

Q1. Please don't use abbreviation in title [CAD]
R1. The title has been changed to: "Micro-calcification detection in mammograms using deep learning"

Q2. what is the AUC in abstract? Area under the curve? Please define
R2. The full term have been added in the revised manuscript:
"area under the curve (AUC)".

Q3. what do the CNN and YOLO stand for in paper? Please define in first usage.
R3. The full terms have been added in the revised manuscript:
"convolutional neural network (CNN)"
"you only look once (YOLO)"

Q4. the authors mentioned "model raining" do they mean "model training?"
R4. This has been corrected in the revised manuscript.

Q5. the authors mentioned: "batch size" do they mean patch size?
R5. No, "batch size" is correct. Batches are used in the training process of CNNs.

Q6. please introduce full term of abbreviations in table 1 in footnote.

R6. Table 1 does not have any abbreviation terms but for Table 2, footnotes have been added in the revised manuscript.

Q7. figure legends have some abbreviations. It could be better if they don't use abbreviation and use full term or introduce these abbreviations.

R7. Full terms have been added in the revised manuscript.

Q8. References 25 to 28 have not been mentioned in paper. Please cite them in paper.

R8. They have been cited in Table 2.

EIC Decision:

Dear authors,

Your article needs Minor Revision.

Regards

5 Revision 4

1 comment · 29 Mar 2022

Editor-in-Chief

ACCEPT 29 Mar 2022

Dear authors,
Your article is acceptable for publication in IJR.
Regards