Journal of Associated Medical Sciences 2024; 57(3): 8-17

JAMS
Journal of Associated Medical Sciences

Scopus Indexed & Thai-Journal Citation Index Centre (TCI)

Journal of Associated Medical Sciences

Journal homepage: https://www.tci-thaijo.org/index.php/bulletinAMS/index

Breast cancer characterization using region-based convolutional neural network with screening and diagnostic mammogram

Jaroonroj Wongnil[1] Anchali Krisanachinda[1] Rajalida Lipikorn[2*]

[1]Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.
[2]Machine Intelligence and Multimedia Information Technology Laboratory, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand.

ARTICLE INFO

Article history:
Received 16 March 2024
Accepted as revised 17 May 2024
Available online 23 May 2024

Keywords:
Malignant calcification, artificial intelligence, two-stage detection model, one-stage detection model, region-based convolutional neural network.

ABSTRACT

Background: Detection and classification of microcalcifications in breast tissues is crucial for early breast cancer diagnosis and long-term treatment.

Objective: This paper aims to propose a robust model capable of detection and classification of breast cancer calcifications in digital mammogram images using Deep Convolutional Neural Networks (DCNN).

Materials and methods: An expert breast radiologist annotated the 3,265 clinical mammogram images to create a comprehensive ground truth dataset comprising 2,500 annotations for malignant and benign calcifications. This dataset was utilized to train our model, a two-stage detection system incorporating a Region-based Convolutional Neural Network (RCNN) with AlexNet and support vector machines to enhance the system's robustness. The proposed model was compared to the one-stage detection, utilizing YOLOv4 combined with the Cross-Stage Partial Darknet53 (CSPDarknet53) architecture. A separate dataset of 504 mammogram images was explicitly set aside for model testing. The efficacy of the proposed model was evaluated based on key performance metrics, including precision, recall, F1 score, and mean average precision (mAP).

Results: The results showed that the proposed RCNN-2 model could automatically identify and categorize calcifications as malignant or benign, outperforming the YOLOv4 models. The RCNN-2's overall effectiveness, as evaluated by precision, recall, F1 score, and mean average precision (mAP), achieved scores of 0.82, 0.85, 0.83, and 0.74, respectively.

Conclusion: The proposed RCNN-2 model demonstrates very effective detection and classification of calcification in mammogram images, especially in high-dense breast images. The performance of the proposed model was compared to that of YOLOv4, and it can be concluded that the proposed RCNN model yields outstanding performance. The model can be a helpful tool for radiologists.

Introduction

Breast cancer is one of the most common cancers and has the highest mortality rate among women worldwide. In 2022, the World Health Organization (WHO) reported that the first rank of new cancer cases in Thai women was breast cancer, with 21,628 patients (23.2%) and the second-highest cause of death in the same year.[1] Mammography is a crucial medical imaging technique and the gold standard for breast cancer detection.[2] Digital mammography, which replaces screen-film technology, provides superior images.[3-5] Its widespread use in breast cancer screening and diagnosis has significantly contributed

* **Corresponding contributor.**
Author's Address: *Machine Intelligence and Multimedia Information Technology Laboratory, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand.*
E-mail address: rajalida.l@chula.ac.th
doi: 10.12982/JAMS.2024.042
E-ISSN: 2539-6056

to early detection, reducing breast cancer mortality by 40%.[6-8] Microcalcification is an early indicator of breast cancer, identified as bright, white spots or dots on the breast tissue in mammography images.[9,10] Malignant calcifications are usually smaller and typically range from <0.5-1 mm.

In contrast, benign calcifications are generally more extensive, with a more defined and coarse appearance.[11-13] However, breast density influences the diagnostic sensitivity and efficacy of mammography.[14,15] The Breast Imaging Reporting and Data System (BI-RADS), developed by the American College of Radiology (ACR), indicates that high-dense breast tissue significantly impacts mammography's sensitivity and diagnostic accuracy.[16-18]

Artificial intelligence powered by deep learning with the convolutional neural network (CNN) has been widely applied in medical imaging.[19] They enable automatic and adaptive feature learning across low to high-level complexity patterns across spatial levels.[20,21] In Thailand, numerous studies have utilized CNNs for the classification of breast cancer in various breast imaging. For instance, Aphinives *et al.* explored AI development using free-trial services to detect microcalcifications in mammography.[22] Additionally, Intasam *et al*. investigated deep learning models for classifying mammograms as benign or malignant.[23] Recently, Labcharoenwongs *et al*. developed an automated breast tumor detection and classification system using deep learning techniques based on the computerized analysis of breast ultrasound images.[24] Despite these advancements, targeted research on detecting and classifying microcalcifications remains relatively limited, especially in high-density breast tissues.

Therefore, this study aims to propose a robust model capable of detecting and characterizing breast cancer calcifications in digital mammogram images using Deep Convolutional Neural Networks (DCNN) by developing a two-stage network architecture utilizing authentic clinical breast images in various formats, coupled with critical parameter adjustments. The model was compared with a one-stage network to determine the most suitable method for detecting and classifying breast cancer calcifications in digital mammography for Thai women.

Materials and methods

This research is a retrospective diagnosis that received approval from the Institutional Review Boards (IRB) of two hospitals that provided digital mammogram images. Due to the retrospective nature, informed consents from patients were waived. Mammogram images from two hospitals were merged to strengthen the learning model's robustness. Patient demographics and pathological data were extracted from electronic medical records (EMR). The process consists of data preparation, model construction, and performance evaluation.

Data preparation

Three thousand two hundred sixty-five clinical mammogram images were collected from January 1, 2018, to December 31, 2019. The dataset comprises Mediolateral Oblique (MLO) and Craniocaudal (CC) views for each breast. Breast density categories, routinely assigned by radiologists in standard clinical workflows using the BI-RADS system, were retrieved from mammography reports.[25,26] Malignant calcifications on mammograms, identified and reviewed by radiologists with histopathological confirmation through biopsy, were also utilized as ground truth. Patients lacking histopathological data were excluded from the research.

The first step was manually removing artifacts from the images, such as location markers and views. Regions of Interest (ROIs) were defined and extracted from MLO and CC views. The ROIs of suspicious areas were manually cropped according to the distribution of calcifications, and adjustments were made to ensure that all relevant areas were included. ROI criteria were derived from radiological and pathological reports executed by expert breast radiologists. Each ROIs was saved as a new image with 227x227 pixels. These ROIs were labeled as one of three categories: malignant calcification, benign calcification, or normal breast tissue. The number of ROIs per image could range from one to four. A total of 5,000 ROI images were generated, consisting of 2,500 ROIs identified as malignant calcifications, 1,250 as benign calcifications, and another 1,250 as normal breast tissue. Examples of ROIs for each category are shown in Figure 1(a).

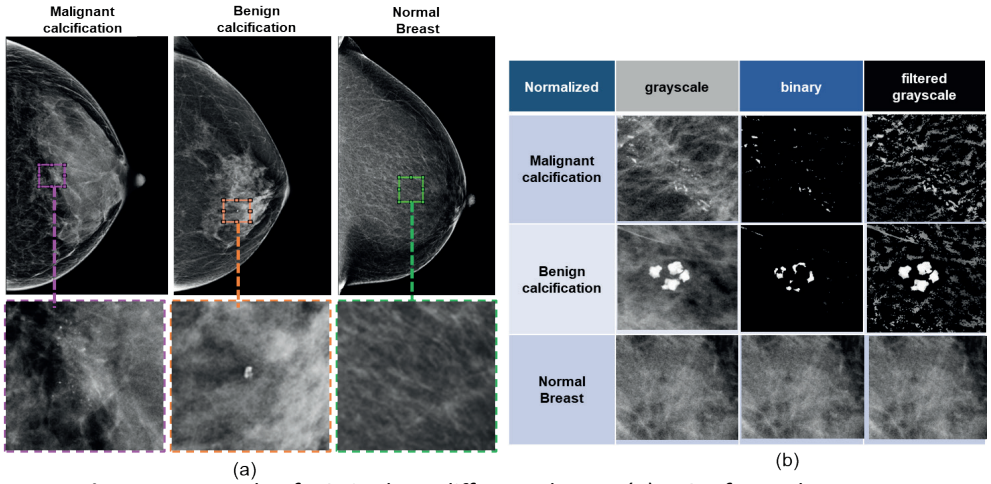


Figure 1. *Example of ROI in three different classes. (a): ROIs for each category, (b): Three sets of generated data from ROIs for training CNN.*

The next step was to generate data from ROIs for training CNN using grayscale normalization, binarization, and logical AND operation to obtain grayscale images, binary images, and filtered images that represent potential real-life scenarios in mammogram imaging, as shown in Figure 1(b). For grayscale normalization, ROI images were transformed and normalized into grayscale with an intensity range between 0 and 1 to ensure consistency and remove variations in intensity scaling, allowing us to focus on structural attributes and intensity fluctuations within the breast tissue that are crucial for detecting abnormalities. For binarization, ROI images were converted to binary images using adaptive thresholding. The foreground polarity was used to specify that the desired foreground (object) is brighter than the background. A threshold value was estimated by setting the sensitivity parameter to 0.5. Any pixels surpassed this threshold value were set to 1 (white), while others were set to 0 (black). These binary images facilitate feature extraction and region-based analysis. For logical operation, the AND operator was applied to each grayscale ROI image and a binary image to produce a filtered grayscale image that highlighted lesions and minimized background distractions. This method helped sharpen lesion visibility, reduces noise, and preserves essential intensity details for better analysis. Consequently, each dataset comprised 5,000 ROI images.

The final stage was to create a ground truth dataset for training and evaluating the proposed and comparative models. To establish the ground truth dataset, 2,500 bounding boxes were drawn on 883 mammogram images displaying malignant features and annotated on 552 images identified with benign features. Each mammogram image may contain up to four bounding boxes. Expert breast radiologists supervised this entire process.

Model construction

The proposed model is a two-stage detection system based on RCNN using AlexNet as the base network, as shown in Figure 2. All computations were performed by MATLAB version R2022a on a personal computer (CPU: Intel Core i7, RAM: 24 GB RAM, NVIDIA 64-bit operating system). A five-fold cross-validation was employed to evaluate and select the model.

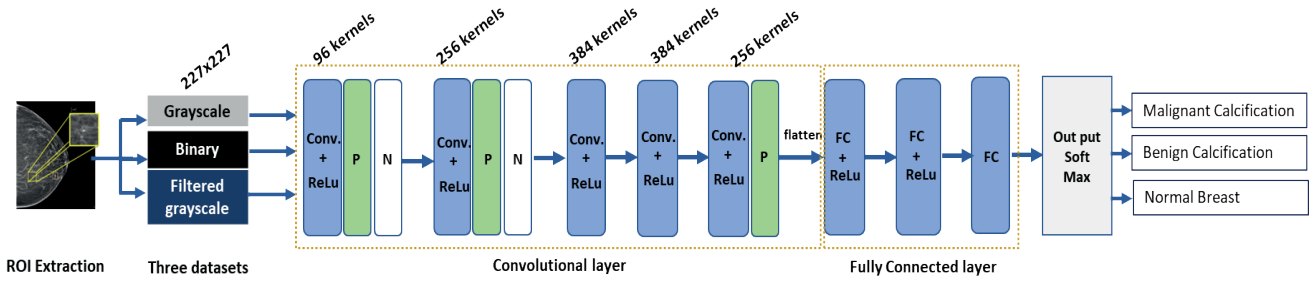


Figure 2. *CNN backbone architecture.*

AlexNet was first trained by 4000 cropped ROI images that were resized to 227x227 pixels for the input layer to improve the classification accuracy. These input data were fed to convolutional layers deploying 96 and 256 kernels for the initial feature map that were enhanced with max pooling and normalization techniques in the first and the second layers. Then, feature analysis was performed through successive convolutional layers, utilizing 384 kernels without pooling until the final convolutional layer. The final convolutional layer was constructed by utilizing 256 kernels. The flattening was used to convert the 2-dimensional arrays obtained from max pooling into a vector. The vector was fed to the fully connected layer. The network's final layer was customized to distinctively craft for three classes: malignant, benign, and normal tissue. The learning rates were increased for quicker adaptation to mammogram data.

A five-fold cross-validation process was employed. AlexNet models 1 through 5 represented each model built from each validation cycle, where the model performances were evaluated from randomly segmented data. The AlexNet model that demonstrated the highest diagnostic precision was chosen as the base network of the RCNN framework, ensuring comprehensive validation against various data patterns and potential anomalies.

Among the 5,000 ROI images in each dataset, 1,000 were used as test data, and the remaining 4,000 ROI images were divided into a training set (80%) and a validation set (20%). The model with the highest performance was selected as the base network of the RCNN model.

The proposed RCNN model for calcification detection and classification that consisted of two processes was constructed starting from using the edge box method[27] to find the region proposals that might contain calcification and using AlexNet as a base network to classify each region proposal whether it is malignant or benign as shown in Figure 3. This figure shows the proposed two-stage model for detecting and classifying breast calcifications. The model begins with an edge box method to locate potential regions of interest (ROIs) or region proposals containing calcifications. These identified region proposals are then processed using AlexNet, which serves as the base network of the proposed RCNN model. The process starts with feature extraction layers, and the sequence involves five convolutional layers (Conv) applying filters to regions of interest (ROIs) to extract detailed features. The output of each layer feeds into the next, while Local Response Normalization (N) is applied after the first and second pooling layers to normalize the responses. Max pooling (P) is applied after the first, second, and fifth convolutional

layers to reduce their dimensionality and processed by ReLU activation functions. The extracted features are flattened and passed through multiple fully connected layers to synthesize the learned information. Finally, a Support Vector Machine (SVM) is integrated to enhance the classification accuracy by optimally separating the identified classes with maximum margins, significantly boosting the precision and robustness of the model. The process culminates in a softmax output layer that classifies the regions into malignant or benign categories, clearly depicting the classification results. The proposed RCNN model was trained with 2,888 images of the whole breast randomly selected from the ground truth dataset. The remaining 377 images of the entire breast were used as test data.

For comparison, the one-stage network model utilizes You Only Look Once version 4 (YOLOv4), known for its efficient one-stage object detection was also implemented. It incorporated the Cross-Stage-Partial-connections Darknet-53 (CSPDarknet53), a 53-layer CNN that used residual connections and Leaky ReLU activation to improve training efficiency and accuracy. YOLOv4 was trained and tested using the same ground truth dataset, ensuring a consistent basis for comparison.

The proposed RCNN model and YOLOv4 were trained to utilize the same ground truth dataset. Critical hyperparameters such as learning rate, batch size, and epochs were meticulously optimized to enhance model performance. This optimization developed distinct configurations: RCNN Models 1 and 2 and YOLOv4 Models 1 and 2. Specifically, RCNN Model 1 and YOLOv4-1 shared hyperparameters with a learning rate 0.001, a batch size of 128, and 50 epochs. Conversely, RCNN Model 2 and YOLOv4-2 were configured with a learning rate 0.0001, maintaining the same batch size but extending to 100 epochs.

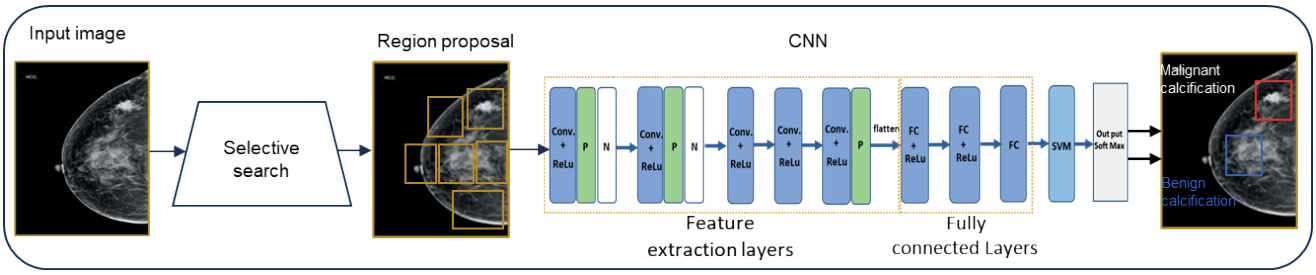


Figure 3. *The proposed two-stage model.*

Performance evaluation

This research utilized multi-statistical metrics to evaluate CNN backbone and the two-stage model and to compare the two-stage and one-stage networks. The equations used in this research are displayed in Table 1. The models' performances were determined using confusion matrices, employing a 3x3 matrix for the multi-class classification by AlexNet and a 2x2 matrix for evaluating the performance of the proposed RCNN model and YOLOv4. These detailed evaluations thoroughly analyzed the models' accuracies and potential utility in medical imaging diagnostics.

Table 1. Confusion matrix of three models.

Evaluation metrics	Equation	Description
True positive	TP	Predicted that a bounding box exists, object is was correct
False positive	FP	Predicted that a bounding box exists, but object is was wrong
False negative	FN	Did not predict a bounding box, even though an object is there
True negative	TN	Not typically defined in the context of object detection evaluation metrics.
Precision (Positive predictive value)	$\frac{TP}{TP+FP} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$	Probability of the predicted bounding boxes that matched the actual ground truth boxes
Recall (Sensitivity, True positive rate)	$\frac{TP}{TP+FN} = \frac{\text{Correct Predictions}}{\text{Total GroundTruth}}$	Probability of correctly detecting ground truth objects
F1 Score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	A balanced performance measure of the model performance (harmonic mean precision and recall)
Micro average F1-Score	$\frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + 1/2(\sum_{i=1}^{n} FP_i + \sum_{i=1}^{n} FN_i)}$	Sums result from all classes, including TP, FN, and FP, to compute an overall F1 score, making it suitable for evaluating models on imbalanced datasets.
Macro average F1 score	$\frac{\sum_{i=1}^{n} F1\_Score_i}{n}$	Averages the F1 scores for all classes, treating each equally, and is ideal for assessing model performance across varied class distributions.
Weighted average F1 score	$\sum_{i=1}^{n} Wi \times F1\_Score_i$	Multiplying each class's F1 score by its proportion in the dataset and then summing these values, providing a metric that accounts for class imbalance.
Mean average precision (mAP)	$\frac{1}{n} \sum_{k=1}^{k=n} APk$	Performance measurement across multiple classes AP_k is AP of class *k*, *n* is the number of classes

Results

The performances of AlexNet across three datasets of 5,000 ROI-cropped images using five-fold cross-validation are shown in Table 2. The results highlight that AlexNet4, AlexNet5, and AlexNet6 achieved the highest accuracies of 90.40%, 87.10%, and 85.00% for the grayscale, binary, and filtered grayscale datasets, respectively. The top-performing model for each dataset was further evaluated on a separate test set of 1,000 ROI images, with the classification outcomes depicted through a 3x3 confusion matrix for each dataset, as shown in Figure 4. Furthermore, the evaluation of the three AlexNet models includes micro average F1 score, macro average F1 score, and weighted average F1 score, as shown in Table 3. Notably, the AlexNet4 model using the grayscale dataset outperforms all other F1 scores, leading to its selection as the base network of the proposed RCNN model.

Table 2. Performances of AlexNet models on three datasets.

Model	Grayscale	Binary	Filtered grayscale
AlexNet1	89.60	85.30	85.00
AlexNet2	88.80	85.30	84.30
AlexNet3	88.80	86.20	83.80
AlexNet4	90.40	84.30	83.30
AlexNet5	89.33	87.10	85.00

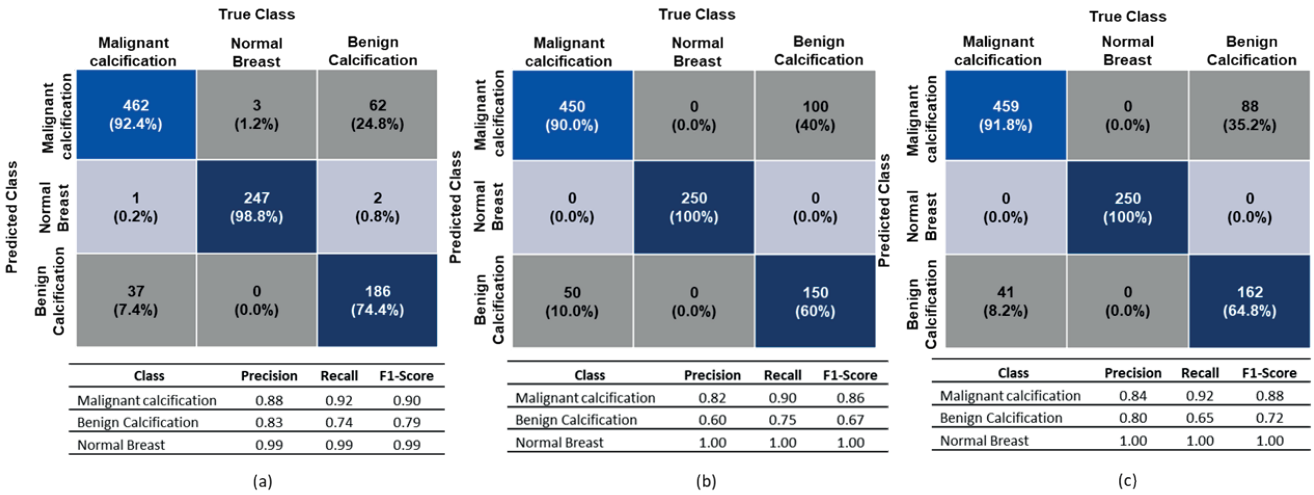


Figure 4. *The 3x3 confusion matrices represent the classification results. (a): AlexNet4 on grayscale ROIs, (b): AlexNet5 on binary ROIs, (c): AlexNet5 on filtered grayscale ROIs with precision, recall, and F1-score for each class.*

Table 3. Performances of the best AlexNet models across three variations of ROI images.

Dataset	Micro average F1	Macro average F1	Weighted F1
AlexNet4 on grayscale	0.90	0.89	0.89
AlexNet5 on binary	0.85	0.84	0.85
AlexNet5 on filtered grayscale	0.87	0.86	0.87

The proposed RCNN models employing AlexNet4 and YOLOv4 with CSPDarknet53 underwent training using the same ground truth dataset. The trainings were conducted under two different configurations to assess the model's effectiveness and the influence of hyperparameters on their performances compared to YOLOv4, as shown in Table 4.

Table 4. Hyperparameters and training time of the detection and classification models.

Detector network model	Learn rate (LR)	Batch size (BS)	Epochs	Training hours
R-CNN-1	0.001	128	50	34.31±3.50
R-CNN-2	0.0001	128	100	66.77±3.53
YOLOv4-1	0.001	128	50	1.87±0.14
YOLOv4-2	0.0001	128	100	3.46+0.14

RCNN-1, RCNN-2, YOLOv4-1, and YOLOv4-2 were performed using a test set of 377 mammogram images. Table 5 summarizes the performances of four models across various metrics, considering both scenarios with a confidence score at a threshold value equal to 0.5 (CF) and without a confidence score (No CF), which provides a holistic view of model performances. Furthermore, the use of five-fold cross-validation and varying confidence score threshold values enable detailed calculations of average precision and recall for each class, enhancing the robustness and clarity of the evaluation.

Table 5. Classification performances of RCNN-1, RCNN-2, YOLOv4-1, and YOLOv4-2 based on precision and recall.

Model	Class		Precision		Recall	
			No CF	CF 0.5	No CF	CF 0.5
R-CNN-1	Malignant calcification	Average	0.31	0.76	0.51	0.58
		SD	0.03	0.04	0.02	0.01
	Benign calcification	Average	0.34	0.75	0.54	0.61
		SD	0.03	0.03	0.01	0.03
R-CNN-2	Malignant calcification	Average	0.57	0.82	0.83	0.84
		SD	0.05	0.03	0.01	0.02
	Benign calcification	Average	0.60	0.83	0.83	0.85
		SD	0.05	0.02	0.02	0.02
YOLOv4-1	Malignant calcification	Average	0.32	0.79	0.53	0.55
		SD	0.03	0.01	0.03	0.01
	Benign calcification	Average	0.34	0.78	0.54	0.56
		SD	0.02	0.03	0.03	0.01
YOLOv4-2	Malignant calcification	Average	0.44	0.73	0.65	*0.75*
		SD	0.05	0.04	0.02	0.03
	Benign calcification	Average	0.43	0.78	0.66	*0.78*
		SD	0.04	0.07	0.02	0.02

Note: No CF: without a confidence factor, CF 0.5: confidence factor at threshold value equal to 0.5.

Table 6 shows the comparative performance metrics of the four models based on average precision, recall, F1-score, and mean Average Precision (mAP). These values are presented along with their ranges to account for variability in the five-fold cross-validation process. The results indicate that RCNN-2 achieves superior performance metrics compared to other models.

Table 6. Performance metrics for the four models.

Model	Precision	Recall	F1 Score	mAP
R-CNN-1	0.72 (0.66-0.78)	0.66 (0.58-0.74)	0.69 (0.65-0.73)	0.66 (0.65-0.67)
R-CNN-2	0.82 (0.80-0.84)	0.85 (0.83-0.87)	0.83 (0.82-0.84)	0.74 (0.73-0.75)
YOLOv4-1	0.72 (0.64-0.80)	0.57 (0.54-0.60)	0.64 (0.54-0.74)	0.70 (0.63-0.77)
YOLOv4-2	0.77 (0.73-0.81)	0.78 (0.74-0.80)	0.77 (0.76-0.78)	0.70 (0.66-0.74)

Discussion

The proposed RCNN model with a grayscale dataset achieved superior performance, as evidenced by the highest F1 scores. As presented in Figure 5, the comparison across different dataset forms clearly shows the advantages of using grayscale images. In the grayscale image (Figure 5a), essential intensity details are maintained, aiding in the more precise differentiation of classes. On the other hand, binary and filtered grayscale images (Figure 5b and c) might simplify the foreground but at the cost of losing fine details, which can be detrimental when analyzing tiny lesions.

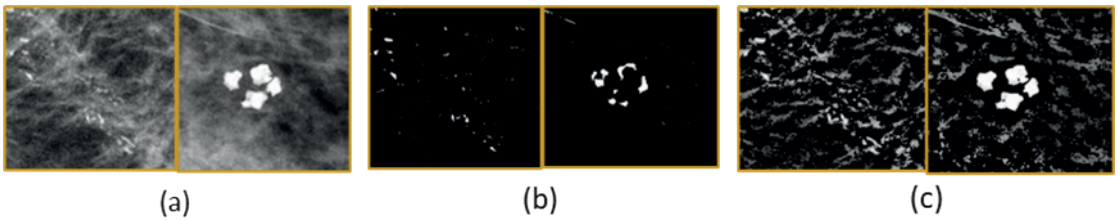


Figure 5. *Images of malignant calcification and benign calcification of breast tissue. (a): grayscale, (b): binary, (c): filtered grayscale.*

This paper proposes the RCNN model that underscores the importance of precise hyperparameter adjustments, notably in learning rates. Such fine-tuning was the key to the RCNN-2 model's exceptional performance, enabling it to analyze complex mammographic patterns accurately. Despite newer CNN models, RCNN was chosen for its proven efficacy in object-scale datasets like mammograms. The RCNN-2 model with a learning rate 0.0001 and 100 epochs yields the highest precision, recall, F1 score, and Map. A low learning rate facilitates a gradual understanding of complex patterns in mammogram images that could lead to better convergence and enhance performance, showcasing its ability to differentiate subtle details in dense breast images and, furthermore, setting the confidence score threshold at 0.5 enhanced detection accuracy across all four models. This threshold level could effectively reduce less reliable detections, particularly in the cases that contributed to partial false positives and false negatives, as depicted in Figure 6. Confidence scoring is pivotal in refining model performance.

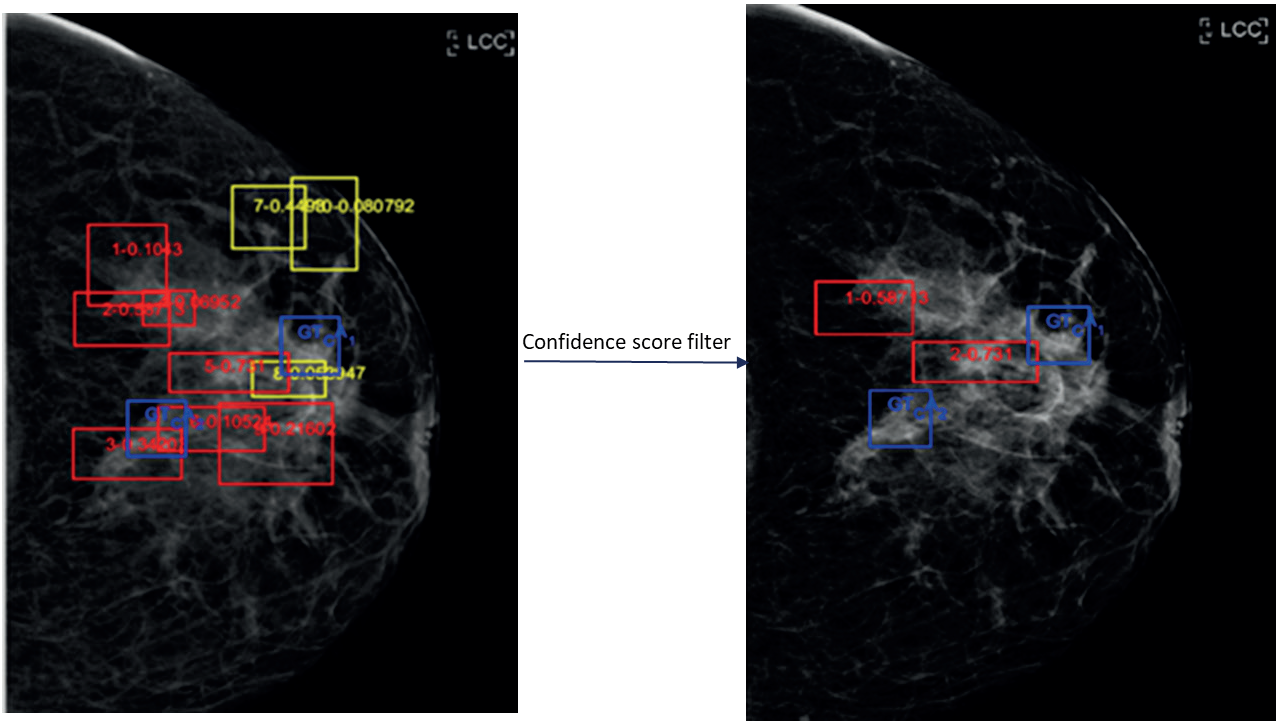
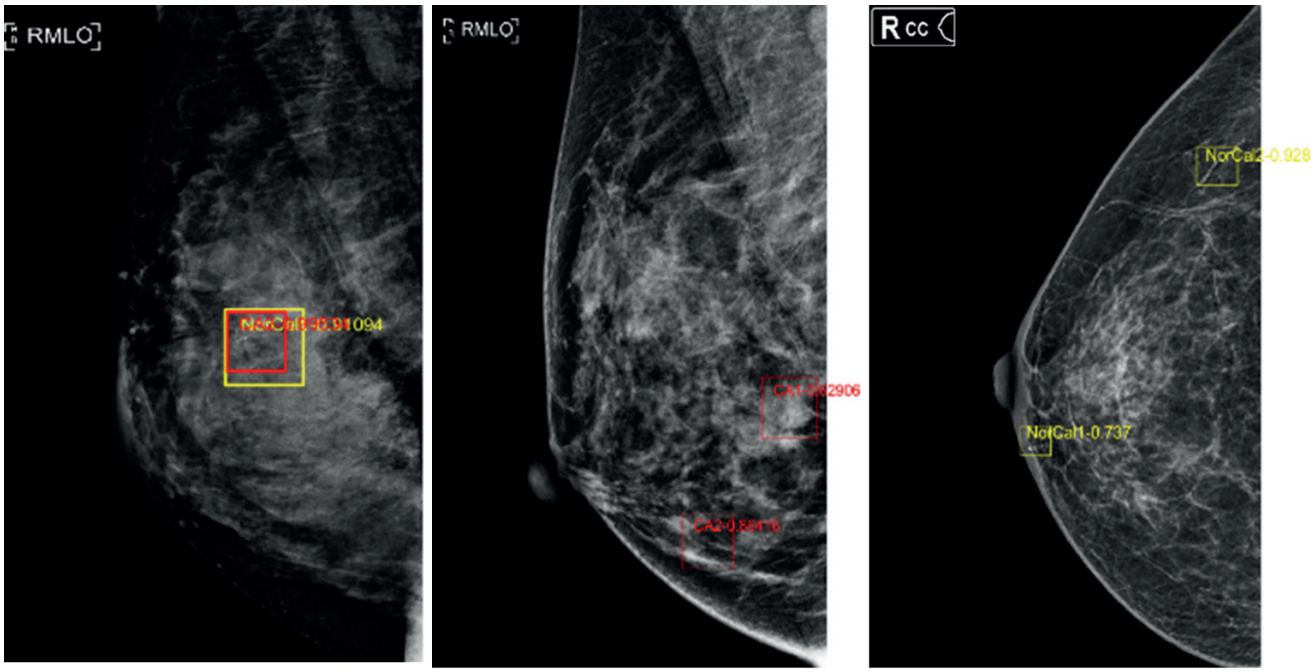


Figure 6. *Utilization of a confidence score helped reduce the impact of unreliable detections that resulted in false positives (FP) and false negatives (FN).*

The RCNN-2 model excels with precision (0.82) and recall (0.85), showcasing its strength in accurate classification and valid positive identification, resulting in an F1 score of 0.83 and mAP of 0.74. Meanwhile, YOLOv4-2 closely follows precision (0.77) and recall (0.78), with an F1 score of 0.77 and mAP of 0.70. Despite lower metrics, RCNN-1 and YOLOv4-1 still post F1 scores of 0.69 and 0.64 with a mAP of 0.70 each. RCNN-2 notably outperforms in detecting and classifying calcifications in dense breasts, a challenging task due to the overlapping characteristics of benign and malignant calcifications.

The proposed model demonstrates its capability to distinguish between benign and malignant microcalcifications in dense breast tissue, as shown in Figure 7(a). This ability indicates that the proposed model can extract and analyze distinctive patterns and characteristics hidden within the highly dense breasts. Figure 7(b) exemplifies highly dense breast tissue accuracy with two correctly classified region proposals. Figure 7(c) highlights the ease of detection in a non-dense breast due to its sharper background.



(a) (b) (c)

Figure 7. *Examples of mammogram images. (a): model's ability to differentiate between benign and malignant microcalcifications in dense breast tissue, (b): two correctly classified region proposals, (c): clear detection in a non-dense breast.*

In clinical validation, the assessment outcomes stem from comparing the model results with the diagnostic results from the radiologist. This section employed the latest mammogram images and preliminary tests to demonstrate the model's effectiveness with current clinical mammogram images. The expert breast radiologists agree with the detection and classification results obtained from the proposed model, demonstrating the model's proficiency in interpreting and classifying mammographic abnormalities. Furthermore, it signifies the model's potential as an auxiliary tool in the diagnostic process. However, it is essential to acknowledge that the images can sometimes lead to disparities in classification. These discrepancies typically arise when the contents of some images present subtle or ambiguous features that require subjective interpretation, especially in high-dense breast mammogram images.

In comparing our research with other studies on breast cancer using deep learning, it is evident that each approach offers unique insights. Aphinives *et al.* highlighted AI's capability in detecting microcalcifications with a precision of 80.0% and a recall of 12.5%, depending on training duration.[22] Intasam *et al*. reported an accuracy of 86.76% by evaluating various CNN architectures.[23] Labcharoenwongs *et al*. advanced a system for tumor detection and volume estimation in ultrasound images, achieving high accuracy and robust classification.[24] Our study employs a two-stage detection system using RCNN integrated with AlexNet, improving robustness and accuracy and addressing both detection and classification of microcalcifications in high-density breast tissues. This comprehensive approach enhances diagnostic tools for early breast cancer detection, especially in challenging dense breast tissues, and has been validated in clinical settings by expert breast radiologists.

Limitation

The limitation of this research is the lack of information on female patients who might be suspected of having breast cancer; it does not consistently offer complete care information, such as pathological reports that affect the collection of mammogram images and the model's ability to fully understand and predict based on localized demographic and clinical nuances. The computational constraints also impact the deep learning model's effectiveness and increase processing times.

Conclusion

In conclusion, this research successfully addresses its objective to develop a robust two-stage Deep Convolutional Neural Network (DCNN) model for detecting and classifying breast cancer calcifications in digital mammogram images. The proposed RCNN model, built upon the proposed CNN based on AlexNet, demonstrates significant potential in aiding radiologists in identifying microcalcifications, particularly within high-density breast tissues of Thai subjects. The comparative analysis with the one-stage YOLOv4 underscored the superior precision and recall scores achieved by our RCNN model, emphasizing the benefits of model fine-tuning and training with varied datasets. Future research will explore other alternative models and develop specifically address the distinct features of Thai female breasts, aiming to enhance the precision and dependability of breast cancer diagnostics. This study also highlights that the identification of breast cancer calcifications in highly dense breasts by expert

radiologists, combined with a model that, while not the latest version, still maintains high capability and is tuned with appropriate parameters, can result in artificial intelligence aiding healthcare professionals in early breast cancer detection. This is particularly crucial in dense breast tissues, where traditional mammography may falter, aiming to improve patient outcomes and screening efficiency.

Conflict of interest

The authors declare no conflict of interest.

Funding

No funding was received to conduct this study.

Ethical Approval

The Ethical Committee of the Faculty of Medicine, Chulalongkorn University, Thailand (IRB no. 503/63), and Bumrungrad International Hospital, Bangkok, Thailand (Grant No. 20:205/CRC:ai) approved the research.

Acknowledgements

The authors would like to thank expert breast radiologists, Department of Radiology, Faculty of Medicine, Chulalongkorn University, and Bumrungrad International Hospital, Bangkok, Thailand, for their guidance and analysis.

References

[1] Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, *et al*. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. 2024. https://gco.iarc.fr/today [accessed 1 February 2024].

[2] Padia K, Douglas TS, Cairncross LL, Baasch RV, Vaughan CL. Detecting Breast Cancer with a Dual-Modality Device. Diagnostics. 2017; 7(1): 17. doi: 10.3390/diagnostics7010017

[3] Haus AG, Yaffe MJ. Screen-Film and Digital Mammography: Image Quality and Radiation Dose Considerations. Radiol Clin North Am. 2000; 38(4): 871-98. doi: 10.1016/s0033-8389(05)70207-4

[4] Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys. 2013; 40(1): 014301. doi: 10.1118/1.4770279

[5] International Atomic Energy Agency, Quality Assurance Programme for Digital Mammography, IAEA Human Health Series No. 17, IAEA, Vienna, 2011.

[6] Yilmaz R, Aydiner A, Igci A, Soran A. Breast Imaging: Breast Cancer A Guide to Clinical Practice. 1st ed. Springer Nature, Switzerland, AG, 2019.

[7] American College of Radiology, Mammography saves lives. https://www.acr.org/Practice-Management-Quality-Informatics/Practice-Toolkit/Patient-Resources/Mammography-Saves-Lives. Accessed on February 20, 2023.

[8] Monticciolo DL, Newell MS, Hendrick RE, Helvie MA, Moy L, Monsees B, *et al*. Breast Cancer Screening for Average-Risk Women: Recommendations From the ACR Commission on Breast Imaging. J Am Coll Radiol. 2017; 14: 1137-43. doi: 10.1016/j.jacr.2017.06.001

[9] Nyante SJ, Lee SS, Benefield T, Hoots TN, Henderson LM. The association between mammographic calcifications and breast cancer prognostic factors in a population-based registry cohort. Cancer. 2017; 123: 219-27. doi: 10.1002/cncr.30281

[10] Logullo AF, Prigenzi KCK, Nimir CCBA, Franco AFV, Campos MSDA. Breast microcalcifications: Past, present and future (Review). Mol Clin Oncol. 2022; 16(4): 81. doi: 10.3892/mco.2022.2514

[11] Cai H, Huang Q, Rong W, Song Y, Li J, Wang J. *et al*. Breast Microcalcification Diagnosis Using Deep Convolutional Neural Network from Digital Mammograms. Comput Math Methods Med. 2019; 1-10. doi: 10.1155/2019/2717454

[12] Chen Z, Strange H, Oliver A, Denton ER, Boggis C, Zwiggelaar R, et al. Topological Modeling and Classification of Mammographic Microcalcification Clusters. IEEE Trans Biomed Eng. 2015; 62: 1203-14. doi: 10.1109/TBME.2014.2385102

[13] Singh N, Marak J, Joshi P, Singh DK. Morphological and distribution pattern of calcifications on full field digital mammography versus digital breast tomosynthesis and comparison of diagnostic abilities of the two modalities: A retrospective study. J Clin of Diagn Res. 2023; 17(3): TC36-TC41. doi: 10.7860/JCDR/2023/55632.17675

[14] Mann RM, Athanasiou A, Baltzer PAT, *et al*. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). Eur Radiol. 2022; 32(6): 4036-45. doi: 10.1007/s00330-022-08617-6

[15] Edmonds CE, O'Brien SR, Conant EF, *et al*. Mammographic Breast Density: Current Assessment Methods, Clinical Implications, and Future Directions. Seminars in Ultrasound, CT, and MRI. Epub. 2022; 44: 35-45. doi: 10.1053/j.sult.2022.11.001

[16] Sickles EA, D'Orsi CJ, Bassett LW. American College of Radiology. ACR BI-RADS® Atlas, 5th ed. Reston, VA, USA, 2013.

[17] Gordon PB. The Impact of Dense Breasts on the Stage of Breast Cancer at Diagnosis: A Review and Options for Supplemental Screening. Curr Oncol. 2022; 17; 29(5): 3595-636. doi: 10.3390/curroncol29050291

[18] Bodewes FTH, van Asselt AA, Dorrius MD, Greuter MJW, de Bock GH. Mammographic breast density and the risk of breast cancer: A systematic review and meta-analysis. Breast. 2022; 66: 62-8. doi: 10.1016/j.breast.2022.09.007

[19] Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evol Intell. 2022; 15(1): 1-22. doi: 10.1007/s12065-020-00540-3

[20] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018; 9; 611-29. doi: 10.1007/s13244-018-0639-9

[21] Ahmed SF, Alam MSB, Hassan M, Rozbu MR, Ishtiak

T, Rafa N, *et al*. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. 2023; 56: 13521-617. doi: 10.1007/s10462-023-10466-8

[22] Aphinives C, Aphinives P, Nawapan S. Artificial Intelligence Development for Detecting Micro-calcification within Mammography. J Med Assoc Thai. 2021; 104(4): 560-4. doi: 10.1002/jmat.2021.104.560

[23] Intasam A, Promworn Y, Thanasitthichai S, Piyawattanametha W. A comparative study of convolutional neural networks for mammogram diagnosis. Proceedings of 14th the Biomedical Engineering International Conference (BMEiCON-2022); 2022 Nov 10-13; Songkhla, Thailand. p.1-4. doi: 10.1109/BMEiCON56653.202210012074. Available from: http://ieeexplore.ieee.org/document/10012074.

[24] Labcharoenwongs P, Vonganansup S, Chunhapran O, Noolek D, Yampaka T. An Automatic Breast Tumor Detection and Classification including Automatic Tumor Volume Estimation Using Deep Learning Technique. Asian Pac J Cancer Prev. 2023; 24(3): 1081-8. doi: 10.31557/APJCP.2023.24.3.1081

[25] Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, *et al*. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. Radiol. 2019; 290(1): 52-8. doi.org/10.1148/radiol.2018180694

[26] Sabani A, Landsmann A, Hejduk P, Schmidt C, Marcon M, Borkowski K, *et al*. BI-RADS-Based Classification of Mammographic Soft Tissue Opacities Using a Deep Convolutional Neural Network. Diagnostics. 2022; 12(7): 1564. doi: 10.3390/diagnostics12071564

[27] Zitnick CL, Dollár P. Edge Boxes: Locating Object Proposals from Edges. European Conference on Computer Vision. 2014: Part V, LNCS 8693 p.391-405.