



OPEN Leveraging paired mammogram views with deep learning for comprehensive breast cancer detection

Jae Won Seo¹, Young Jae Kim² & Kwang Gi Kim^{1,3}✉

Employing two standard mammography views is crucial for radiologists, providing comprehensive insights for reliable clinical evaluations. This study introduces paired mammogram view based-network (PMVnet), a novel algorithm designed to enhance breast lesion detection by integrating relational information from paired whole mammograms, addressing the limitations of current methods. Utilizing 1,636 private mammograms, PMVnet combines cosine similarity and the squeeze-and-excitation method within a U-shaped architecture to leverage correlated information. Performance comparisons with single view-based models with VGGnet16, Resnet50, and EfficientnetB5 as encoders revealed PMVnet's superior capability. Using VGGnet16, PMVnet achieved a Dice Similarity Coefficient (DSC) of 0.709 in segmentation and a recall of 0.950 at 0.156 false positives per image (FPPI) in detection tasks, outperforming the single-view model, which had a DSC of 0.579 and a recall of 0.813 at 0.188 FPPI. These findings demonstrate PMVnet's effectiveness in reducing false positives and avoiding missed true positives, suggesting its potential as a practical tool in computer-aided diagnosis systems. PMVnet can significantly enhance breast lesion detection, aiding radiologists in making more precise evaluations and improving patient outcomes. Future applications of PMVnet may offer substantial benefits in clinical settings, improving patient care through enhanced diagnostic accuracy.

Keywords Mammogram, Breast, Artificial intelligence, Detection, Computer-aided diagnostic system

Background

Breast cancer is a formidable challenge in the realm of public health, significantly influencing mortality statistics worldwide¹. However, from a clinical perspective, approximately 90% of cancers can be cured if detected and treated at an early stage, thus emphasizing the importance of early diagnosis of breast cancer². Detecting breast lesions, which are the principal indicators of breast cancer, during the nascent stages is crucial for improving survival probabilities and the efficacy of therapeutic interventions. Therefore, there is a continuous demand for improving the accuracy of breast lesion diagnosis based on high sensitivity and low false positive rates³. Mammography is an indispensable primary diagnostic modality in breast cancer screening, providing multiple images depending on the angle of view⁴. Among the various views, the craniocaudal view (CC view), obtained by compressing the breast up and down, and the mediolateral oblique view (MLO view), obtained by adjusting the angle to include the pectoralis muscles, are used as standard views for screening^{5,6}. The use of these two views in the clinical interpretation of patients with suspected breast cancer has been reported to be clinically advantageous in reducing false positives (FPs) and improving sensitivity compared with using only a single view^{7–10}.

In the contemporary era, advances in artificial intelligence (AI) have led to remarkable progress in medical imaging as well as computer vision. This has significantly reduced the burden on clinicians by accelerating the diagnosis process. Moreover, AI-based CAD has proven its worth by providing consistent and objective indicators, demonstrating its reliability and efficiency. This is evident in its positive impact on inter- and intra-observer^{11–14}. Deep learning-based convolutional neural networks (CNN), a field of artificial intelligence, extract features as filter maps based on convolutional operations. This approach, which preserves the location

¹Department of Health Sciences and Technology, GAIHST, Gachon University, Incheon 21999, Republic of Korea.

²Department of Gachon Biomedical & Convergence Institute, Gachon University Gil Medical Center, Incheon 21565, Republic of Korea. ³Department of Biomedical Engineering, College of IT Convergence, Gachon University, Seongnam-si 13120, Republic of Korea. ✉email: kimkg@gachon.ac.kr

information of the extracted feature values, has found widespread use in computer vision and has demonstrated excellent efficiency in medical image data^{15,16}. As a result, it has become a key tool in the development of breast lesion detection algorithms for early diagnosis of breast cancer from mammography data.

Related works

AI algorithms for breast lesion detection have been mainly based on single images. The Faster R-CNN model is an object detection algorithm that sequentially divides region detection and detected region classification based on convolutional neural networks¹⁷. This model has been used variously through application and structural modification to detect the location of breast lumps with relatively high detection accuracy and to classify lesions as benign or malignant^{18–20}. In addition, the YOLO model, which has the advantage of fast inference speed while simultaneously detecting the location of candidate regions for the target object and classifying the detected regions, has been used mainly by using the model structure as it is or by modifying its layers^{21–23}. The Retinanet model, proposing to improve the loss function to train about small objects efficiently, has been utilized in various breast lesion detection studies²⁴. However, these studies were single view-based algorithms, where the lesion detection was performed separately for each image, and the correlated information between the two reference images, which could potentially enhance the accuracy of the detection, was not considered.

Some studies have utilized patches for breast lesion regions to apply the matching information of the two standard images. There is an experiment that extracted feature maps through a convolutional layer for the manually extracted patches from each image and performed patch-by-patch classification of whether the patches are for the same lesion by measuring the similarity of the neighboring values for each pixel in the two feature maps²⁵. Some studies used deep learning-based detection models to extract patches for the lesion region individually and then similarly classified the two patches' matching. One study utilized Unet to extract individual patches from two images. For each patch, a feature map was extracted using a convolutional filter, and each feature map was combined in a fully connected layer to share the features of the two images²⁶. In addition, another study extracted patches using the YOLO model. It calculated cosine similarity between the two feature maps extracted through the convolutional layer from the patches to classify the matching²⁷. These experiments aimed to remove false positives by determining the matching at the patch level. In this case, the positional information of the whole image was not considered in the feature extraction process for matching between patches. Nevertheless, when the two standard images employed the similarity of the lesion patches, they showed improved results in terms of quantitative performance indicators, demonstrating the validity of utilizing both images.

Another way to enhance performance has been conducted by using multi-view mammograms. One study attempted to employ the two views using the distance between the nipple and lesion and considering the location information from the CC view; however, in these studies, the MLO view image was not applied to detect breast lesions and was only used as an auxiliary input for detection in the CC view²⁸. Other studies attempted to reconcile the two or four multi-view mammograms using an ensemble network across whole feature maps. However, they only employed the multi-views simultaneously for classification problems (abnormality or benign and malignant), not for extracting the feature map when detecting mass^{29–31}.

Contributions

Therefore, in this study, we aimed to improve the detection performance by proposing an AI lesion detection algorithm that utilizes paired standard mammograms. The key points of the proposed contributions in this paper are as follows:

1. We propose a deep learning based PMVnet (paired mammogram view based-network) that simultaneously utilizes the whole mammogram image's paired craniocaudal and mediolateral oblique views.
2. To evaluate the proposed algorithm, we compared its performance with single mammogram view-based networks(SMVnet), which are original standard Unet-based models.
3. We analyzed localization results to investigate whether the PMVnet effectively extracts feature maps correlated with two views.

Materials and methods

Data collection

Due to the retrospective nature of the study, the Institutional Review Board (IRB) of Gachon University Gil Medical Center (IRB Number: GCIRB2020-477) waived the need for obtaining informed consent. All experimental protocols were performed in accordance with the relevant guidelines and regulations in compliance with the Declaration of Helsinki. We collected data acquired from devices produced by two different manufacturers: Lorad Selenia 2D Digital Mammography model, referred to as "Hologic" in this paper, from Hologic Inc. based in Lesionachusetts, USA; Senographe Essential model, referred to as "GE" in this paper, from General Electric Healthcare based in Illinois, USA. Figure 1 shows an example of the four standard views (RCC, LCC, RMLO, and LMLO) obtained from a single patient. To simultaneously utilize paired images, we grouped two standard views (RCC and RMLO) for the right side and two standard views (LCC and LMLO) for the left side into a unified set. The dataset comprises 818 sets of left and right views (1,636 images), with 525 sets (1,050 images) sourced from Hologic devices and 293 sets (586 images) obtained from GE. The Hologic data had a resolution of $2,560 \times 3,328$ pixels, whereas GE data had a resolution of $3,328 \times 4,096$ pixels. Of the entire dataset, 738 cases (1,476 images) were allocated to train the models, and the remaining 80 cases (160 images) served as the evaluation data to validate the performance. In the training dataset, the average age of patients was 51.255 years (± 11.651), with 99.6% female and 0.4% male. For performance assessment, the patients were all female and had an average age of 54.263 years (± 11.392). All identifiable information was anonymized for privacy purposes, except for age

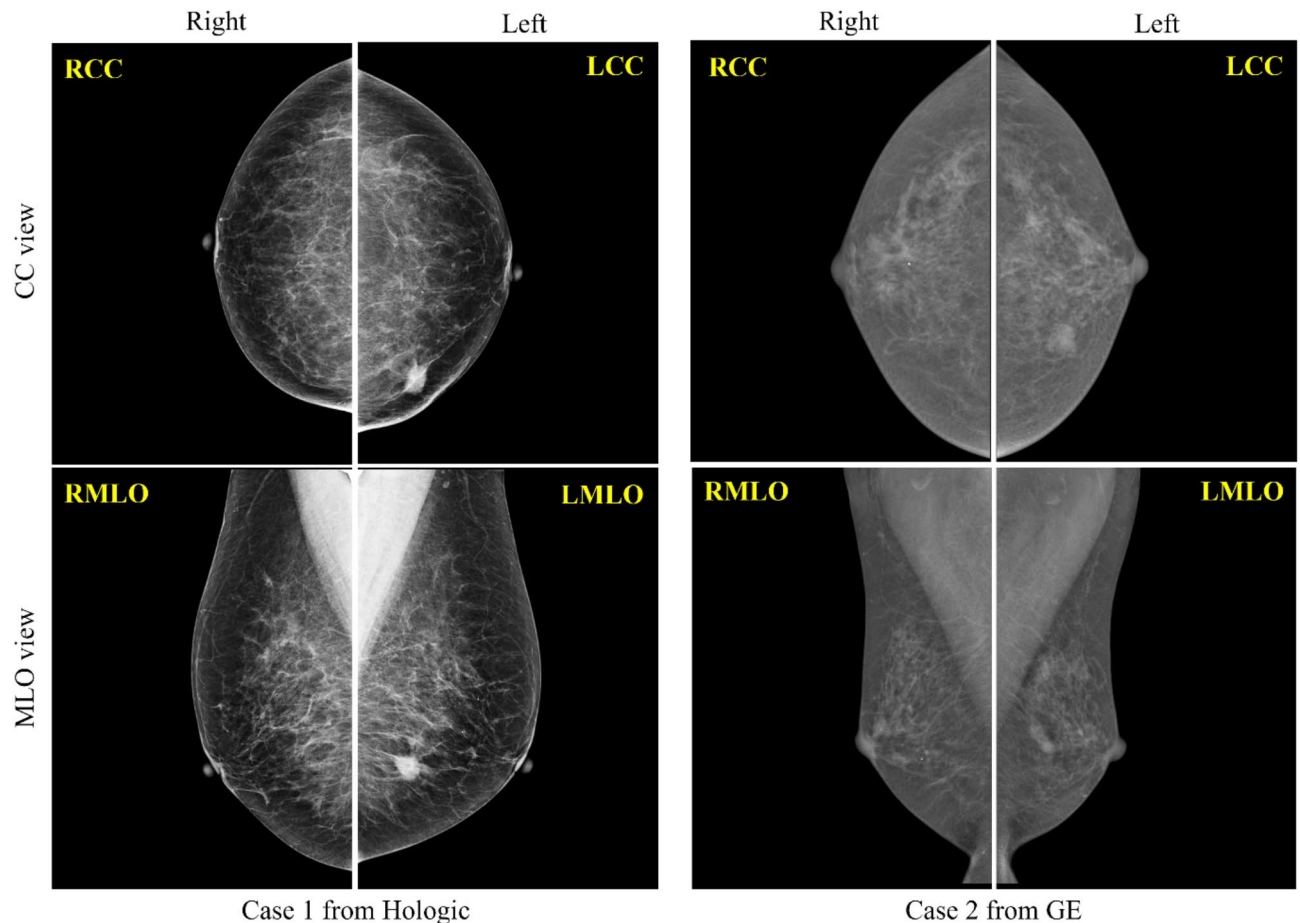


Fig. 1. Examples of standard mammogram views in the left and right directions. *LCC* left craniocaudal view, *LMLO* left mediolateral oblique view, *RCC* right craniocaudal view, *RMLO* right mediolateral oblique view.

and sex. All the data used in this research consisted of cases diagnosed as abnormal (malignant and/or benign lesions).

Data preprocessing

Breast lesions have vague outlines when they overlap with fibroglandular tissue, particularly in high-density breast tissue^{32,33}. To address this issue, various studies have demonstrated that using pseudo-color images as a preprocessing technique can improve contrast and enhance the detection performance of masses, particularly in high-density breast tissue^{34–36}. Therefore, we generated pseudo-color images as a preprocessing step to increase the contrast of the surrounding breast lesions, as shown in Fig. 2. Using pixel intensity, the window settings were adjusted in three different manners. Three adjusted grayscale images were assigned to each of the three channels (channels 0, 1, and 2), and synthesized to create a pseudo-colored image. In channel 0, an image set to the default window width and window level at the time of acquisition using the machine's internal algorithm was inserted. Channel 1 was set to the window width and level for the range of the maximum and minimum values of the pixel value distribution for the breast region. Finally, Channel 2 was set by adjusting the pixel values by setting the window level to the most pixel value in the breast region.

We undertook a resizing process to enhance the utility of the collected mammogram images, which possess resolutions of $2,560 \times 3,328$ pixels or $3,328 \times 4,096$ pixels. The high resolution of these images posed potential challenges related to computational demand and memory usage. As a solution, we reduced the height of the images to 512 pixels, with the width adjusted correspondingly to maintain the original aspect ratio through zero-padding, resulting in a final image size of 512×512 pixels. Furthermore, acknowledging the critical relationship between the volume of training data and model performance, we applied data augmentation techniques to enrich our dataset. By utilizing the padded images, we resized them to ratios of 4/3 and 2 times the original reference image. This approach enabled us to generate additional variations through both horizontal and vertical transitions, effectively increasing our dataset fourfold. An example of the data augmentation methodology is presented in Fig. 3.

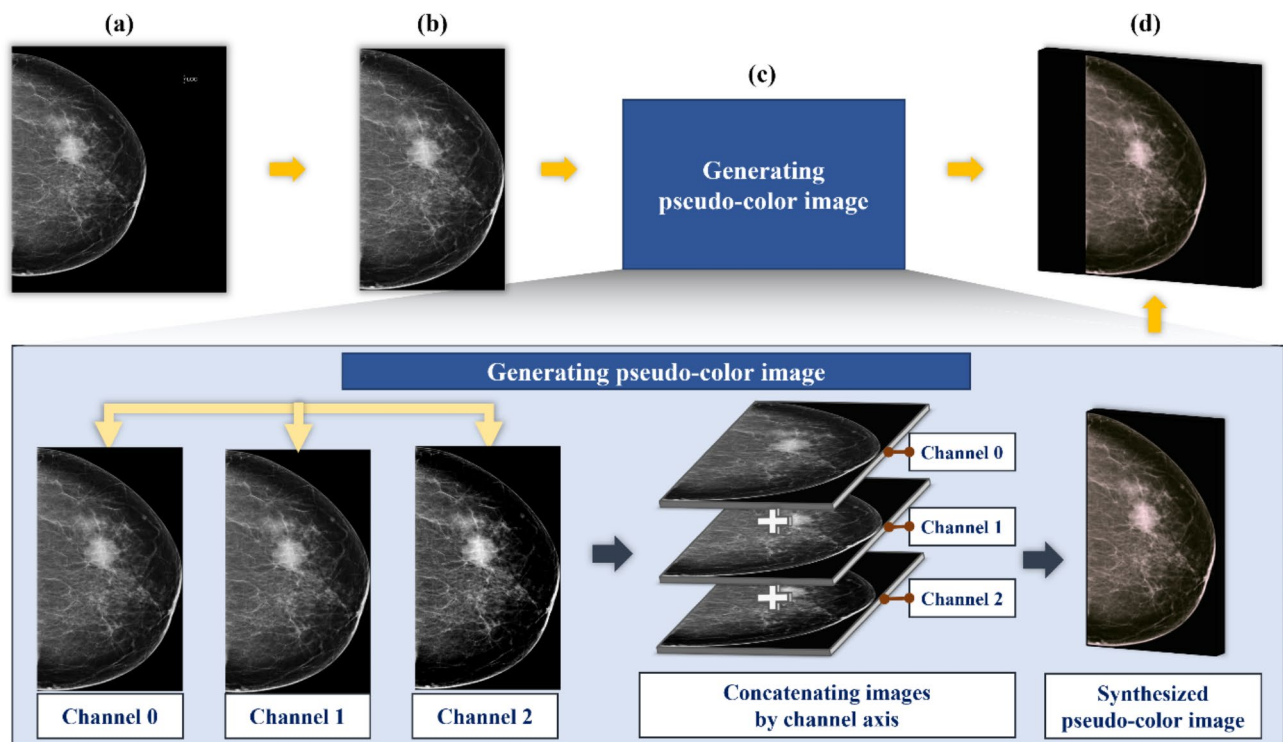


Fig. 2. Example of pseudo-color image generation process to enhance the contrast between the lesion and surrounding area. (a) Original image. (b) The denoised and background-removed breast region. (c) The pseudo-color image generation process and the synthesized pseudo-color image. (d) The input image with padding and scaling for training a network.

Paired standard mammogram views network

The Unet is a prominent CNN-based semantic segmentation model widely applied in medical image segmentation due to its exceptional performance, as noted across various studies^{37–40}. This research introduces a PMVnet structure with a Unet comprising two core components: a contracting path (encoder) and an expanding path (decoder) with a multi-input design⁴¹.

To clearly understand the baseline model used for comparison, we define SMVnet as a single-view mammography network. Unlike PMVnet, which utilizes paired views to improve performance, SMVnet is a traditional Unet-based model that processes only one mammography view at a time. The SMVnet model in this study leverages the Unet architecture with various encoders such as VGG, ResNet, and Efficientnet to ensure a fair and thorough performance evaluation. By comparing the SMVnet-based model with PMVnet, we aim to directly evaluate the benefits of integrating paired-view information.

Figure 4 shows an example of the proposed model structure. This approach involves sharing weights by treating two standard images as a unified representation of one breast. We employed cosine similarity and an attention mechanism based on squeeze and excitation (SE) techniques⁴², effectively integrating the feature maps between the two images. The models trained on an Nvidia V100 GPU with configurations as follows: 4 batch sizes; 100 epochs; learning rate of 0.0001; focal loss; the learning rate was set to decrease by a factor of 0.1 if the loss did not decrease for 15 epochs on the validation data.

Let the outputs extracted from the encoders of each view be O_{cc} and O_{mlo} , respectively, and the concatenation of the two outputs be O . Let the size of the feature map of the O be $H \times W \times C$, the c^{th} feature map by F_c , squeezing the spatial information be S_c , the output applied the SE technique to the O be O' , the operation by the fully connected layer with n outputs be $C_n(x)$, the sigmoid function be $\sigma(x)$ and the Relu function be $R(x)$, and the following operations are applied. In this study, the reduction ratio was set to 4.

$$S_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_c \quad (1)$$

$$O' = \sigma \left(C_n \left(R \left(C_{\frac{n}{4}} (S_c) \right) \right) \right) \quad (2)$$

Therefore, we let the feature maps activated by the SE technique be O'_{cc} and O'_{mlo} , respectively, they can be represented as follows:

$$O'_{cc} = O_{cc} \cdot O' \quad (3)$$

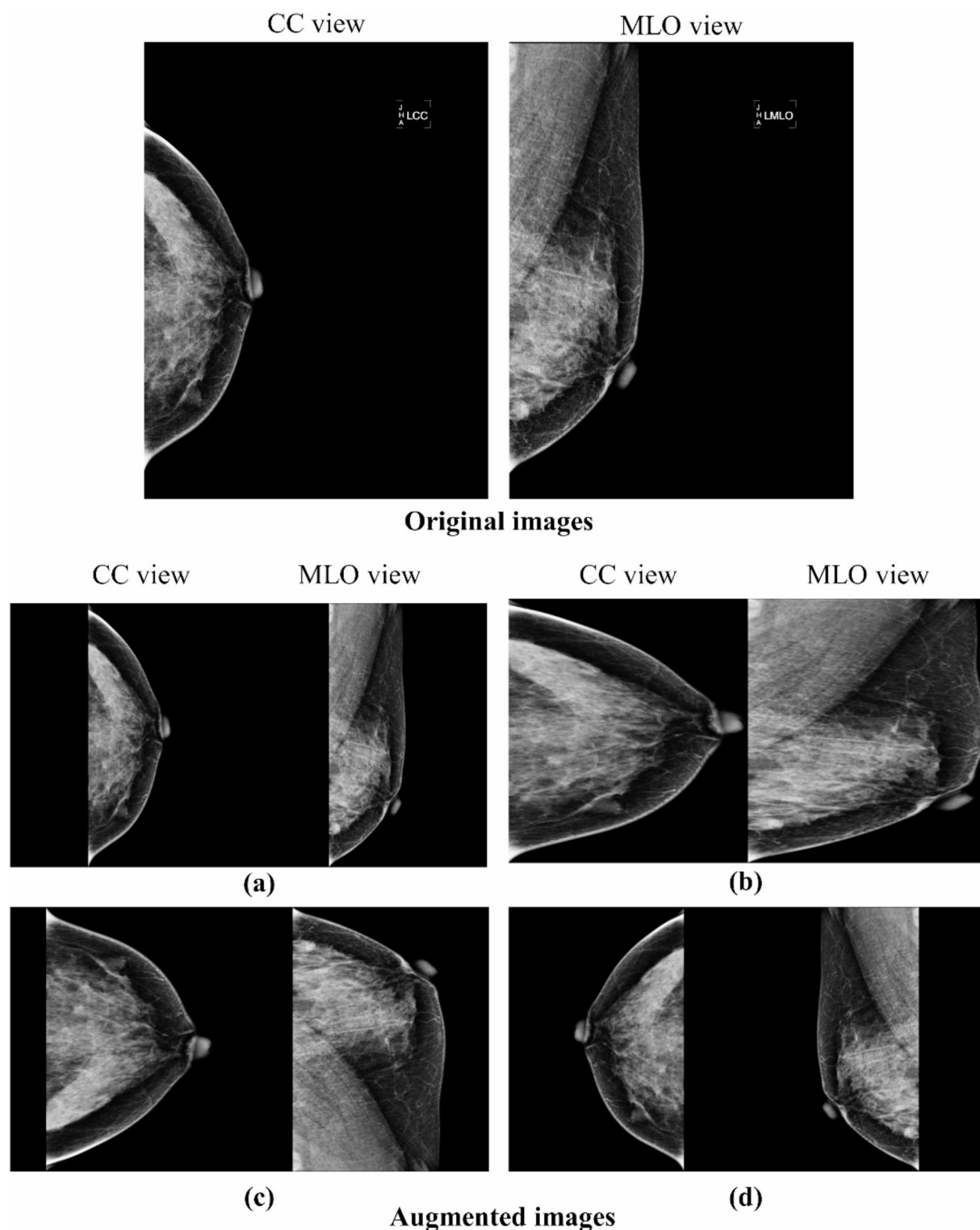


Fig. 3. Examples of augmented images with the data scaling process applied. The final preprocessed image was zero-padded to be 512×512 in size. (a) Resized by keeping the original aspect ratio centered on the center of the image and applying padding on either side. (b) Resized to 512×512 , ignoring the original proportions. (c) Resized to $4/3$ times the original aspect ratio and flipped up and down. (d) Resized to 2 times the original aspect ratio and flipped left and right.

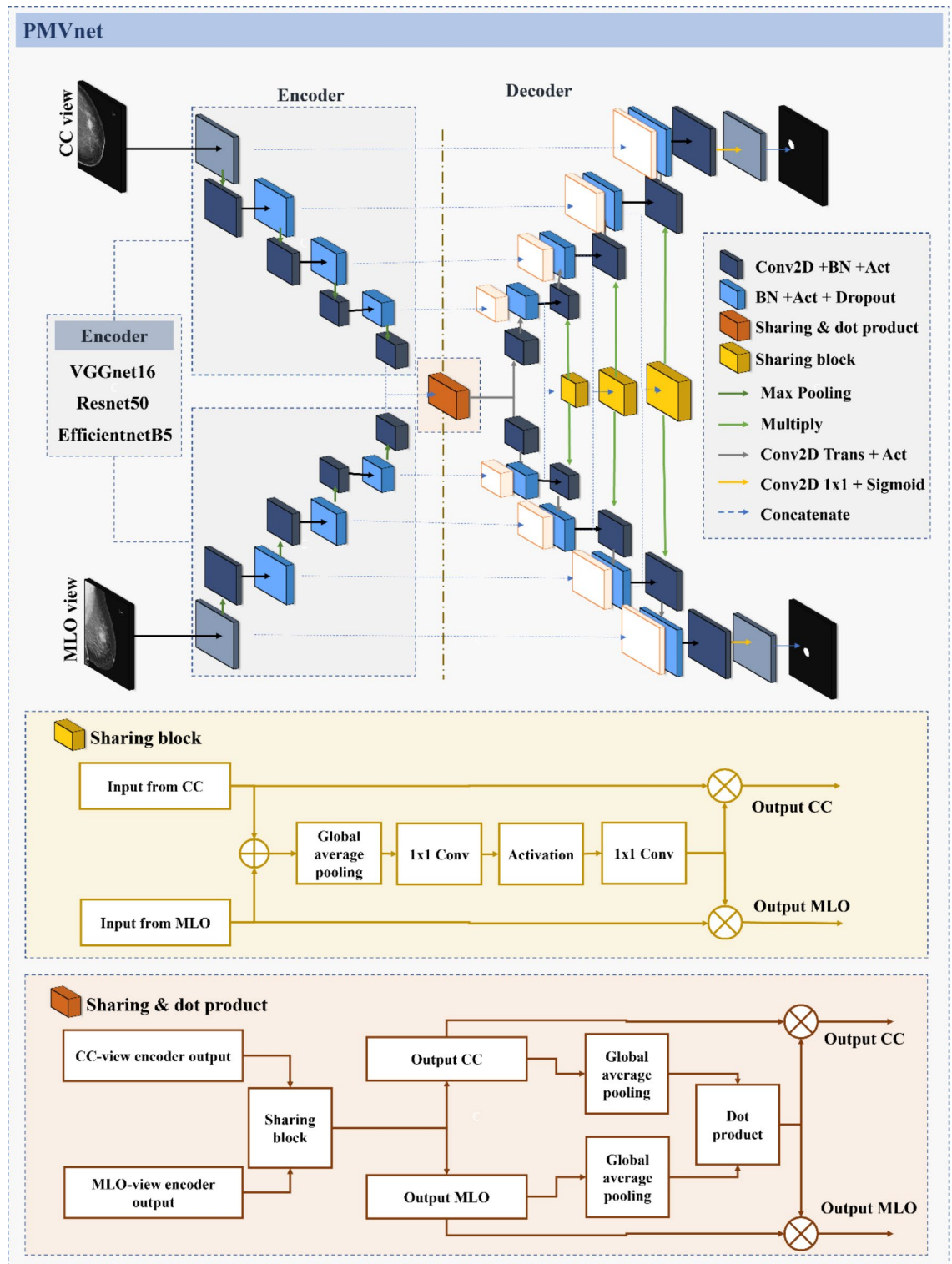


Fig. 4. The proposed PMVnet structure connects two U-shaped based models to form multiple inputs, and the feature maps extracted through each CC and MLO encoder are combined and fused, and then the squeeze-and-excitation method is applied to each feature map at the decoder to measure the cosine similarity based on dot product.

$$O'_{mlo} = O_{mlo} \cdot O' \quad (4)$$

Equation (5) is applied to the results extracted using Eqs. (3) and (4), and the result is applied to the feature map for each image through a scalar product to ensure that the similarity between the two images contributes to the feature extraction.

$$f_{\cos}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (5)$$

Performance assessment

The performances of the models were evaluated and compared to verify their effectiveness. The data used for the performance evaluation were 80 mammograms of the CC view and 80 mammograms of the MLO view that were not used for training. The performance of each model was evaluated using the same data. TP indicates that the model predicted a lesion area as a lesion area, FP indicates that the model predicted a no-lesion area as a lesion area, and FN indicates that the model predicted a lesion area as not a lesion area. TN indicates that the model predicted a no-lesion area as not a lesion area. The TP, FP, FN, and TN were calculated by comparing the ground truth generated by the radiologist. The breast lesion area was predicted by each model on a pixel-by-pixel basis for semantic segmentation assessment using dice similarity coefficient (DSC), recall (same as sensitivity), and precision.

To evaluate lesion detection performance for investigating localization results, we generated a bounding box that included the lesion area (Fig. 5). The intersection of union (IoU) is an evaluation metric based on the overlap between two regions. In this study, we measured the IoU for the ground truth and the predicted region using a model based on the region containing the lesion and defined the measured IoU value for the predicted region as TP if it was above 0.1, FP if it was below 0.1, and FN if there was no predicted region to compare with the reference image region. Using these metrics, we calculated the recall, precision, and false positives per image (FPPI) to derive and analyze the performance of the model in detecting lesion-containing regions. The evaluation formulae are as follows:

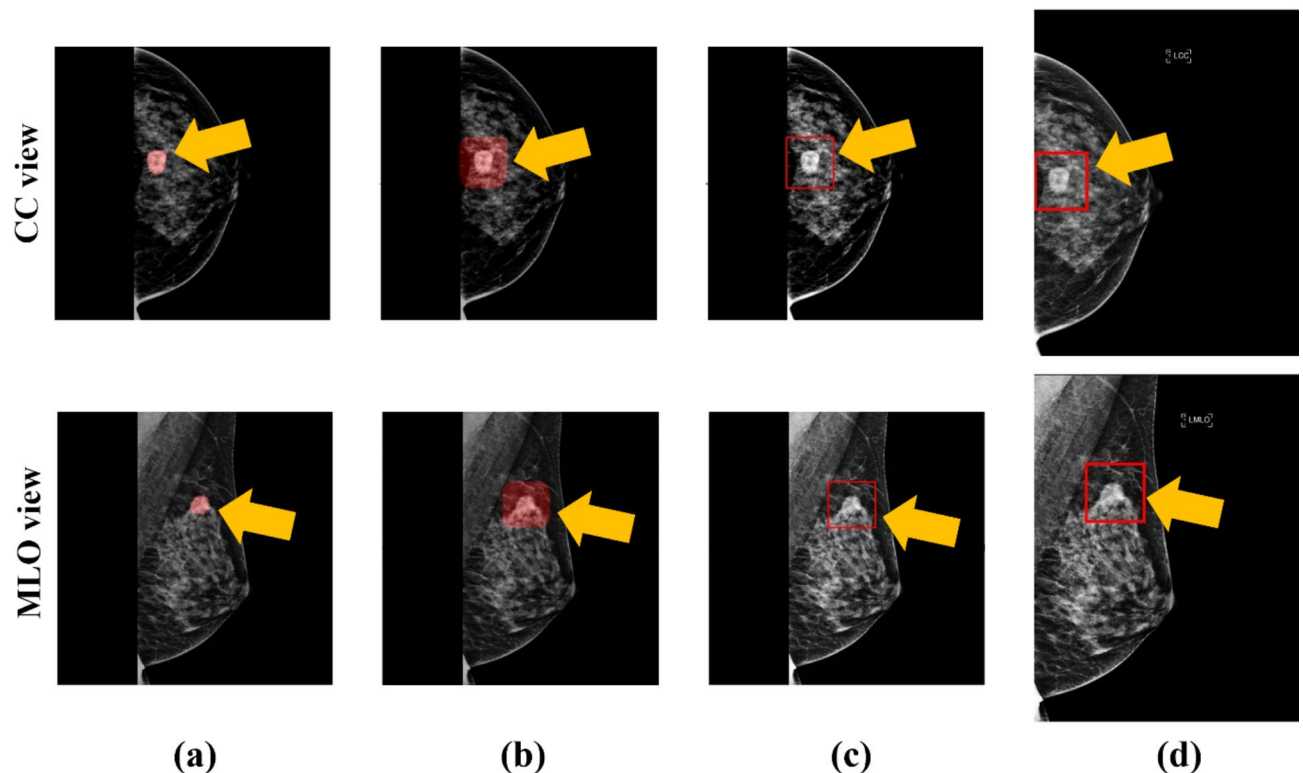


Fig. 5. Example of generating lesion regions for detection performance evaluation from semantic segmentation results. In (c) and (d), the red rectangle represents the converted rectangular format of the detection result containing the lesion. (a) Segmented lesions extracted from the model. The red-blended areas represent lesion regions predicted using the proposed model. (b) The converted area includes lesions generated by morphological operations from the segmentation results. (c) Conversion to the lesion detection format. The red squares represent the lesion-containing regions detected by the morphology operation based on the model predictions. (d) the final format of the lesion region reconstructed from the image size changes during preprocessing, considering the proportion and position of the original image.

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$FPPI = \frac{FP}{the\ Number\ of\ images} \tag{9}$$

Results

In this study, VGGnet16, Resnet50, and EfficientnetB5 were employed as encoder networks in the proposed U-shaped SMVnet and PMVnet-based models. The study aimed to assess the effectiveness of the proposed models and understand the underlying reasons by evaluating the performance of each configuration. In addition, a pretrained model with ImageNet weights was applied to the encoder structure of each model, and the results were compared⁴³.

Table 1 lists the performance evaluation results for semantic segmentation, detailing how variations in the encoding stage of each model influenced the outcomes, as observed across different experiments. In the models without pretrained weights (plain), the proposed PMVnet resulted in a DSC of 0.709 for the VGGnet16 encoder-based model, 0.652 for the Resnet50-based model, and 0.707 for the EfficientnetB5 model. Except for based on EfficientnetB5, the PMVnet-based model showed higher performance in all the values. However, when the models were initialized by ImageNet weights, overall values in PMVnet outperformed SMVnet. Figure 6 illustrates the semantic segmentation outcomes of both SMVnet and PMVnet with frameworks that employ VGGnet16, Resnet50, and EfficientnetB5 models as the encoder.

Table 2 compares the detection performance of each model for the region containing the lesion. It shows that the VGGnet16-based model achieved an overall performance improvement of 0.137 for recall and 0.032 for FPPI, and the Resnet50-based model achieved an improvement of 0.012 for recall and 0.050 for FPPI for the PMVnet. The EfficientnetB5-based model showed an improvement in sensitivity of 0.013 and a decrease in FPPI of 0.013. In the pixel-to-pixel comparison of semantic segmentation, the EfficientnetB5-based paired view model performed worse than the single-view model across all performance metrics but better in sensitivity in the region containing the lesion.

Discussion

To investigate the effectiveness of PMVnet in improving the performance of breast lesion detection, we analyzed the performance assessment results for segmentation and detection with and without ImageNet weights. The segmentation results show that the VGGnet16-based encoders exhibit the largest performance improvement when using PMVnet, followed by Resnet50 and EfficientnetB5, and that the performance improvement decreases as the model structure becomes more complex. PMVnet, designed to extract individual feature maps for each view image and concatenate the feature maps, experiences a rapid increase in the number of parameters as the encoder structure deepens and becomes more complex. Due to this parameter proliferation, applying PMVnet to the same architecture without structural modifications may impede model optimization. Furthermore, the highest performance of VGGnet16, the shallowest and simplest structure in this study, indicates that it is optimized for extracting feature maps from data owing to its structural characteristics. Therefore, in this study, we show that the model size and depth are not necessarily correlated to the performance of the networks and data. Numerous studies indicate no direct correlation between model size and performance. To enhance performance, various strategies have been employed, including modifying the activation function, refining the loss function, and adjusting factors such as the resolution of the input image, the depth of the model layers, and

ImageNet weight	Model	DSC (95% CI)	Precision (95% CI)	Recall (95% CI)
w/o.	VGGnet16 based-PMVnet	0.709 (0.671–0.748)	0.693 (0.652–0.734)	0.782 (0.742–0.821)
	VGGnet16 based-SMVnet	0.579 (0.531–0.627)	0.590 (0.542–0.639)	0.653 (0.601–0.705)
	Resnet50 based-PMVnet	0.652 (0.608–0.696)	0.638 (0.592–0.683)	0.712 (0.665–0.758)
	Resnet50 based-SMVnet	0.643 (0.598–0.688)	0.624 (0.578–0.670)	0.730 (0.682–0.778)
	EfficientnetB5 based-PMVnet	0.707 (0.666–0.748)	0.692 (0.651–0.734)	0.765 (0.722–0.808)
	EfficientnetB5 based-SMVnet	0.709 (0.669–0.750)	0.688 (0.646–0.730)	0.792 (0.750–0.833)
w.	VGGnet16 based-PMVnet	0.690 (0.648–0.732)	0.687 (0.643–0.731)	0.739 (0.695–0.783)
	VGGnet16 based-SMVnet	0.644 (0.599–0.689)	0.655 (0.609–0.701)	0.699 (0.651–0.747)
	Resnet50 based-PMVnet	0.677 (0.635–0.720)	0.676 (0.632–0.719)	0.734 (0.689–0.779)
	Resnet50 based-SMVnet	0.658 (0.614–0.703)	0.658 (0.613–0.703)	0.729 (0.682–0.776)
	EfficientnetB5 based-PMVnet	0.743 (0.707–0.779)	0.725 (0.686–0.763)	0.801 (0.763–0.839)
	EfficientnetB5 based-SMVnet	0.733 (0.695–0.772)	0.715 (0.675–0.756)	0.810 (0.772–0.849)

Table 1. Performance results of semantic segmentation. w with, w/o. without.

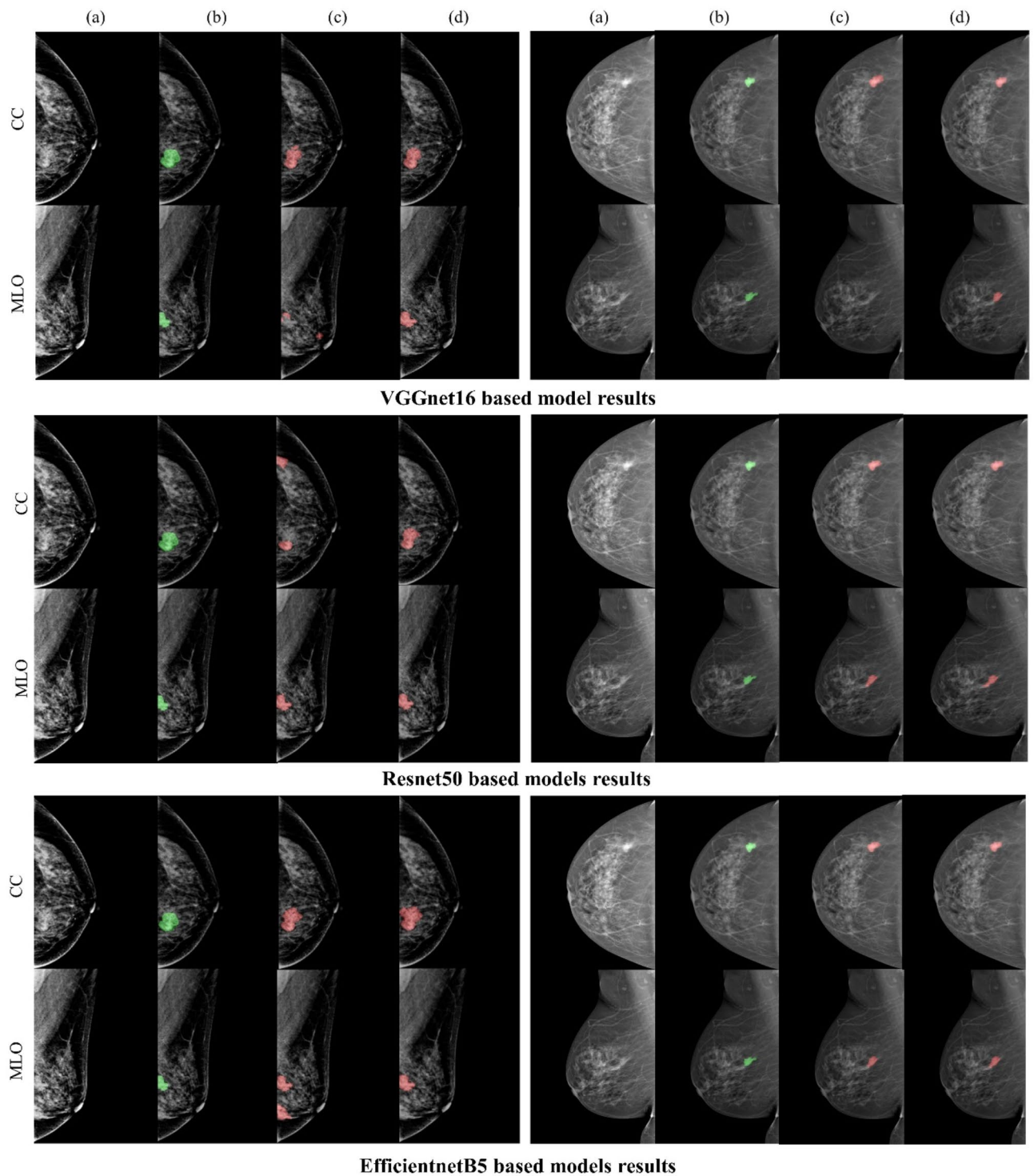


Fig. 6. Example of segmentation results from models utilizing encoders based on VGGnet16, Resnet50, and EfficientnetB5 for the Hologic (left) and GE (right) cases. (a) Original image. (b) Ground truth. (c) SMVnet-based model results. (d) PMVnet-based model results.

the number of channels in the extracted feature map^{44,45}. In the future, we must develop a model with an optimal architecture by considering components such as the size of the input data and the number of channel layers.

Moreover, to verify the suitability of the PMVnet for sharing significant positional information between the two standard views, we analyzed the detection performance by converting it into breast lesion inclusion regions. For all models, the recall value of the detection performance was proportional to the DSC value of the segmentation result; however, in contrast to the segmentation performance of the EfficientnetB5 model, the model with PMVnet showed higher performance. Figure 7 shows examples of conversion to breast lesion areas

ImageNet weight	Model	Recall (95% CI)	Precision (95% CI)	FPPI (95% CI)
w/o.	VGGnet16 based-PMVnet	0.950 (0.912–0.988)	0.900 (0.859–0.941)	0.156 (0.092–0.220)
	VGGnet16 based-SMVnet	0.813 (0.749–0.876)	0.747 (0.684–0.809)	0.188 (0.118–0.257)
	Resnet50 based-PMVnet	0.900 (0.854–0.946)	0.868 (0.819–0.916)	0.175 (0.102–0.248)
	Resnet50 based-SMVnet	0.888 (0.839–0.936)	0.834 (0.783–0.886)	0.225 (0.149–0.301)
	EfficientnetB5 based-PMVnet	0.919 (0.874–0.964)	0.906 (0.863–0.950)	0.113 (0.060–0.165)
	EfficientnetB5 based-SMVnet	0.906 (0.862–0.951)	0.906 (0.862–0.951)	0.100 (0.053–0.147)
w.	VGGnet16 based-PMVnet	0.906 (0.862–0.951)	0.884 (0.838–0.931)	0.144 (0.087–0.201)
	VGGnet16 based-SMVnet	0.856 (0.802–0.911)	0.793 (0.736–0.849)	0.181 (0.119–0.244)
	Resnet50 based-PMVnet	0.922 (0.877–0.967)	0.884 (0.839–0.930)	0.156 (0.090–0.223)
	Resnet50 based-SMVnet	0.888 (0.839–0.936)	0.853 (0.803–0.903)	0.188 (0.111–0.264)
	EfficientnetB5 based-PMVnet	0.931 (0.892–0.971)	0.903 (0.861–0.945)	0.131 (0.076–0.187)
	EfficientnetB5 based-SMVnet	0.938 (0.901–0.974)	0.909 (0.870–0.949)	0.131 (0.073–0.189)

Table 2. Performance results of lesion detection. *w* with, *w/o.* without.

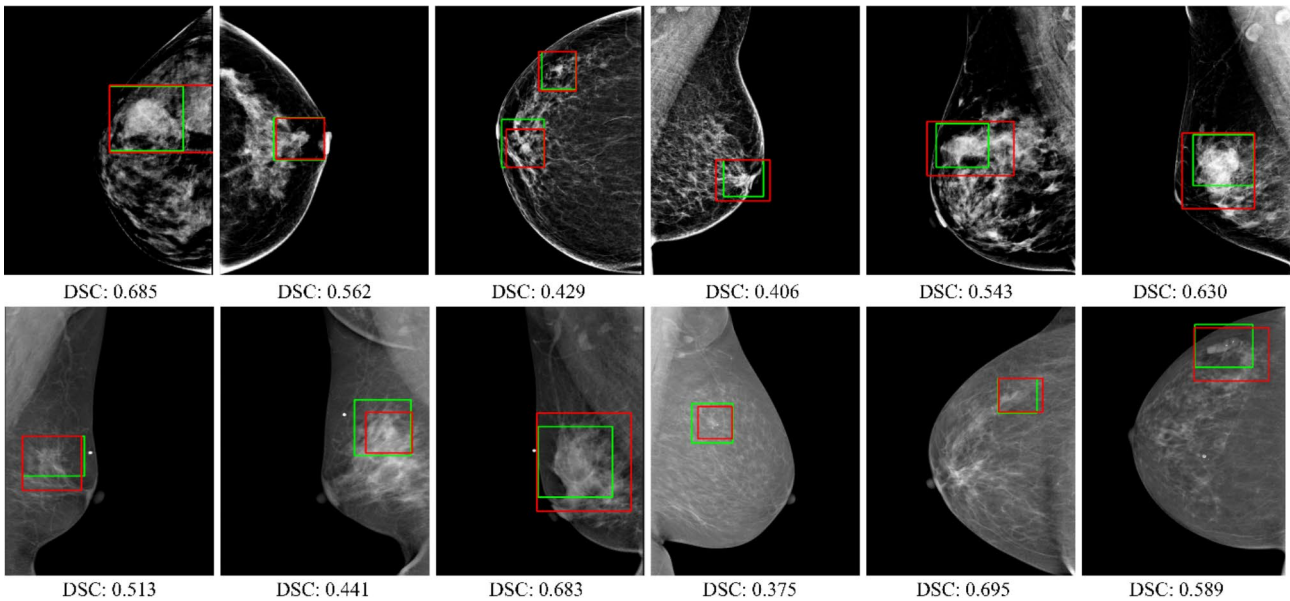


Fig. 7. Example of lesion-containing region detection results from a PMVnet-based model utilizing the EfficientB5 encoder, where the Dice similarity coefficient (DSC) is less than or equal to 0.707 (average) but is rated as a true positive (TP) in the detection region performance evaluation. Red squares represent regions the model estimates; green squares represent ground truth.

by the EfficientnetB5-based PMVnet model, which showed a low DSC performance below 0.7; however, the detection performance was measured as TP. The results demonstrated that the large size of the model and the excessive number of parameters in PMVnet did not significantly affect the extraction of feature maps required for details in pixels but were significant in extracting effective feature maps for localization. Additionally, the lower segmentation performance compared to detection is attributed to the challenges in accurately predicting the boundaries and shape of the breast lesion.

Breast lesions on mammography can be subdivided according to shape, margin, and density according to the Breast Imaging Reporting and Data System (BI-RADS), and there are significant differences between each type. However, in this study, the type of breast lesion type was not considered when constructing the training data, and all lesion types were classified as a single lesion. Thus, the prediction of the center of the lesion, which shows a relatively high contrast for most lesion types in mammograms, was successful; however, the prediction of the boundaries of various lesion types was not. The distinction between benign and malignant lesions is crucial in diagnosing breast cancer, and the type of lesion is a significant indicator of this distinction. Therefore, in future studies, a higher accuracy performance can be achieved by subdividing the types and developing data distribution and learning models that consider them. Moreover, the classification of benign and malignant lesions is an essential process in breast cancer diagnosis, and it is expected that the development of additional models for the classification of such indicators will be crucial in assisting breast cancer diagnosis.

Except for VGGnet16, utilizing ImageNet weights resulted in an overall improvement in the segmentation and detection performance compared with the unweighted cases. Using ImageNet weights hindered feature map extraction in the best-optimized model for the medical data used in this study by initializing the weights fitted in real-world images. The detection performance results mirrored the improvement pattern seen in segmentation. However, models based on EfficientnetB5 exhibited lower performance in Precision and FPPI for detection, indicating that initializing to ImageNet did not universally enhance performance across all models in this study. This result indicates that a comparative study on initializing the optimal weights and the optimized model structure is necessary in the future. In addition, because breast cancer diagnosis uses mammograms as well as breast ultrasound and magnetic resonance imaging data, we plan to develop PMVNet-based algorithms optimized for breast lesion detection by fusing data from various modalities.

While this study presents a novel approach for effectively merging dual-view mammographic information through the proposed PMVNet, certain limitations exist. First, since this study mainly focused on the structural and methodological aspects of integrating paired mammography views to improve diagnostic accuracy, we did not analyze the computational efficiency. This lack of processing time and efficiency is important for real-world clinical applications. In future work, we will evaluate and optimize the computational efficiency of the proposed method to increase its effectiveness in real-world settings. In addition, the dataset used in this study consisted of only abnormal cases, which limits the generalizability to datasets containing both normal and abnormal cases. In future studies, incorporating a balanced dataset will help ensure the proposed model's broad applicability and robustness. Finally, although this study's performance analysis was objectively validated based on quantitative evaluation, it did not include subjective evaluation by expert radiologists, which affects real-world usability. Incorporating expert feedback in future studies would provide valuable insights into the proposed method's clinical validity and utility. Furthermore, the effectiveness of PMVNet needs to be validated in more extensive and diverse populations to ensure its applicability in real-world clinical settings. Testing the model across varied demographic groups will help assess its generalizability and robustness in diverse patient populations, which is an essential next step for ensuring the clinical utility of the proposed method.

Conclusion

In this study, we proposed a PMVNet, paired mammogram view images-based breast mass detection algorithm, considering that both CC and MLO views are utilized as reference images in clinical diagnosis. The PMVNet applied the SE method and cosine similarity to the feature maps extracted from the two view images to share the weights so that the feature information of the two images can be employed by each other. To verify PMVNet, we compared it to the performance of the single-view-based algorithm, and the results showed overall improved detection performance compared to the contrast algorithm. However, as the model's size increased, the proposed algorithm's effectiveness decreased, and the Efficientnet-based model showed the same or even relatively lower performance values. This is because many more parameters are extracted from the PMVNet structure, and the more complex encoder structure does not help improve performance due to excessive parameters unsuitable for fitting the data. In future work, we will develop a structured algorithm to optimize it. In addition, applying the clinical diagnostic approach utilized in this study to mammograms and 3D-based digital breast tomosynthesis data is expected to help clinicians diagnose breast cancer as a computer-aided diagnosis system.

Data availability

The data used to support the findings of this study are available upon request from the corresponding author.

Received: 7 July 2024; Accepted: 31 January 2025

Published online: 05 February 2025

References

1. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *Ca-a Cancer J. Clin.* **73**, 17–48. <https://doi.org/10.3322/caac.21763> (2023).
2. Katanoda, K. & Matsuda, T. Five-year relative survival rate of breast Cancer in the USA, Europe and Japan. *Jpn J. Clin. Oncol.* **44**, 611–611. <https://doi.org/10.1093/jjco/hyu073> (2014).
3. Li, Y., Chen, H., Cao, L. & Ma, J. A survey of computer-aided detection of breast cancer with mammography. *J. Health Med. Inf.* **4**, 6 (2016).
4. Nam, K. J. et al. Comparison of full-field digital mammography and digital breast tomosynthesis in ultrasonography-detected breast cancers. *Breast* **24**, 649–655. <https://doi.org/10.1016/j.breast.2015.07.039> (2015).
5. Majid, A. S., de Paredes, E. S., Doherty, R. D., Sharma, N. R. & Salvador, X. Missed breast carcinoma: pitfalls and pearls. *Radiographics* **23**, 881–895. <https://doi.org/10.1148/rq.234025083> (2003).
6. Popli, M. B., Teotia, R., Narang, M. & Krishna, H. Breast positioning during mammography: mistakes to be avoided. *Breast cancer: basic and clinical research* **8**, BCBCR. S17617 (2014).
7. Jorgensen, K. J. & Bewley, S. Breast-Cancer screening—viewpoint of the IARC Working Group. *N Engl. J. Med.* **373**, 1478. <https://doi.org/10.1056/NEJMc1508733> (2015).
8. Vijayarajan, S. & Jaganathan, P. Breast cancer segmentation and detection using multi-view mammogram. *Acad. J. Cancer Res.* **7**, 131–140 (2014).
9. Warren, R. M., Duffy, S. W. & Bashir, S. The value of the second view in screening mammography. *Br. J. Radiol.* **69**, 105–108. <https://doi.org/10.1259/0007-1285-69-818-105> (1996).
10. Hackshaw, A. K., Wald, N. J., Michell, M. J., Field, S. & Wilson, A. R. An investigation into why two-view mammography is better than one-view in breast cancer screening. *Clin. Radiol.* **55**, 454–458. <https://doi.org/10.1053/crad.2000.0448> (2000).
11. Najjar, R. Redefining radiology: a review of artificial intelligence integration in medical imaging. *Diagnostics* **13**, 2760 (2023).
12. Wenderott, K., Gambashidze, N. & Weigl, M. Integration of Artificial Intelligence Into Sociotechnical Work Systems-Effects of Artificial Intelligence Solutions in Medical Imaging on Clinical Efficiency: protocol for a systematic literature review. *Jmir Res. Protocols*. **11** <https://doi.org/10.2196/40485> (2022).

13. Naeem, O. B. et al. Breast mammograms diagnosis using deep learning: state of art Tutorial Review. *Arch. Comput. Methods Eng.*, 1–19 (2024).
14. Sahu, A., Das, P. K. & Meher, S. An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomed. Signal Process. Control.* **87**, 105377 (2024).
15. Muduli, D., Dash, R. & Majhi, B. Automated diagnosis of breast cancer using multi-modal datasets: a deep convolution neural network based approach. *Biomed. Signal Process. Control.* **71** <https://doi.org/10.1016/j.bspc.2021.102825> (2022).
16. Abdelhafiz, D., Yang, C., Ammar, R. & Nabavi, S. Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinform.* **20**, 281. <https://doi.org/10.1186/s12859-019-2823-4> (2019).
17. Ren, S. Q., He, K. M., Girshick, R., Sun, J. & Faster, R-C-N-N. Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149. <https://doi.org/10.1109/Tpami.2016.2577031> (2017).
18. Akselrod-Balling, A. et al. A CNN based method for automatic mass detection and classification in mammograms. *Comput. Methods Biomech. Biomedical Engineering-Imaging Visualization.* **7**, 242–249. <https://doi.org/10.1080/21681163.2017.1350206> (2019).
19. Ribli, D., Horváth, A., Unger, Z., Pollner, P. & Csabai, I. Detecting and classifying lesions in mammograms with deep learning. *Sci. Rep.* **8** <https://doi.org/10.1038/s41598-018-22437-z> (2018).
20. Agarwal, R., Diaz, O., Yap, M. H., Lladó, X. & Martí, R. Deep learning for mass detection in full field Digital mammograms. *Comput. Biol. Med.* **121** <https://doi.org/10.1016/j.compbiomed.2020.103774> (2020).
21. Al-masni, M. A. et al. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Comput. Methods Programs Biomed.* **157**, 85–94. <https://doi.org/10.1016/j.cmpb.2018.01.017> (2018).
22. Baccouche, A., Garcia-Zapirain, B., Olea, C. C. & Elmaghraby, A. S. Breast lesions detection and classification via YOLO-Based Fusion models. *Cmc-Computers Mater. Continua.* **69**, 1407–1425. <https://doi.org/10.32604/cmc.2021.018461> (2021).
23. Al-masni, M. A. et al. Detection and Classification of the Breast Abnormalities in Digital Mammograms via Regional Convolutional Neural Network. *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Embc)*, 1230–1233 (2017). (2017).
24. Jung, H. et al. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. *Plos One.* **13** <https://doi.org/10.1371/journal.pone.0203355> (2018).
25. AlGhamdi, M. & Abdel-Mottaleb, M. D. V. D. C. N. N. Dual-view deep convolutional neural network for matching detected masses in mammograms. *Comput. Methods Programs Biomed.* **207**, 106152. <https://doi.org/10.1016/j.cmpb.2021.106152> (2021).
26. Perek, S., Hazan, A., Barkan, E. & Akselrod-Ballin, A. Siamese network for dual-view Mammography Mass Matching. *Image Anal. Mov. Organ. Breast Thorac. Images.* **11040**, 55–63. https://doi.org/10.1007/978-3-030-00946-5_6 (2018).
27. Yan, Y. et al. Towards improved breast mass detection using dual-view mammogram matching. *Med. Image Anal.* **71**, 102083. <https://doi.org/10.1016/j.media.2021.102083> (2021).
28. Yang, Z. et al. MommiNet-v2: mammographic multi-view mass identification networks. *Med. Image Anal.* **73**, 102204. <https://doi.org/10.1016/j.media.2021.102204> (2021).
29. Kebede, S. R. et al. Dual view deep learning for enhanced breast cancer screening using mammography. *Sci. Rep.* **14** <https://doi.org/10.1038/s41598-023-50797-8> (2024).
30. Carneiro, G., Nascimento, J. & Bradley, A. P. in *International conference on medical image computing and computer-assisted intervention*. 652–660 (Springer).
31. Wu, N. et al. Deep neural networks improve radiologists' performance in breast Cancer screening. *IEEE Trans. Med. Imaging.* **39**, 1184–1194. <https://doi.org/10.1109/TMI.2019.2945514> (2020).
32. DH, K., Choi, S. & Kim, D. The image quality according to target-filter combination in digital mammography. *J. Korean Soc. Radiological Technol.* **9** (2015).
33. Lee, S. H. et al. Glandular tissue component and breast Cancer risk in Mammographically dense breasts at Screening Breast US. *Radiology* **301**, 57–65. <https://doi.org/10.1148/radiol.2021210367> (2021).
34. Haindl, M. & Remeš, V. Pseudocolor enhancement of mammogram texture abnormalities. *Mach. Vis. Appl.* **30**, 785–794 (2019).
35. Jones, M. A., Faiz, R., Qiu, Y. & Zheng, B. Improving mammography lesion classification by optimal fusion of handcrafted and deep transfer learning features. *Phys. Med. Biol.* **67** <https://doi.org/10.1088/1361-6560/ac5297> (2022).
36. Min, H. et al. in *2020 IEEE 17th International symposium on biomedical imaging (ISBI)*. 1111–1115 (IEEE).
37. Siddique, N., Paheding, S., Elkin, C. P. & Devabhaktuni, V. U-Net and its variants for medical image segmentation: a review of theory and applications. *Ieee Access.* **9**, 82031–82057. <https://doi.org/10.1109/Access.2021.3086020> (2021).
38. Weng, Y., Zhou, T. B., Li, Y. J. & Qiu, X. Y. NAS-Unet: neural Architecture Search for Medical Image Segmentation. *Ieee Access.* **7**, 44247–44257. <https://doi.org/10.1109/Access.2019.2908991> (2019).
39. Liu, X. B., Song, L. P., Liu, S. & Zhang, Y. D. A review of Deep-Learning-Based Medical Image Segmentation methods. *Sustainability* **13** <https://doi.org/10.3390/su13031224> (2021).
40. Liu, L. L. et al. A survey on U-shaped networks in medical image segmentations. *Neurocomputing* **409**, 244–258. <https://doi.org/10.1016/j.neucom.2020.05.070> (2020).
41. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Med. Image Comput. Computer-Assisted Intervention Pt Iii.* **9351**, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28 (2015).
42. Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372> (2020).
43. Deng, J. et al. ImageNet: A Large-Scale Hierarchical Image Database. *Cvpr: Ieee Conference on Computer Vision and Pattern Recognition, Vols 1–4*, 248–255. (2009). <https://doi.org/10.1109/cvpr.2009.5206848> (2009).
44. Tan, M. X. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning, Vol 97* 97 (2019).
45. Pandey, S., Chen, K. F. & Dam, E. B. Comprehensive Multimodal Segmentation in Medical Imaging: combining YOLOv8 with SAM and HQ-SAM models. *2023 Ieee/Cvf Int. Conf. Comput. Vis. Workshops Iccvw.* **2584–2590** <https://doi.org/10.1109/Iccvw60793.2023.00273> (2023).

Acknowledgements

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711137885, KMDF_PR_20200901_0012-02), the GRRC program of Gyeonggi Province [GRRC-Gachon2023(B01)], and the Gachon Program (GCU-202307640001).

Author contributions

Jae Won Seo: Conceptualization, Data curation, Methodology, Visualization, Formal analysis, Investigation, Writing - original draft. Young Jae Kim: Conceptualization, Investigation, Validation, Project administration, Writing - review & editing. Kwang Gi Kim: Validation, Funding acquisition, Supervision, Writing - review & editing.

Declarations

Competing interests

The authors declare no competing interests.

Ethical standards

The Institutional Review Board (IRB) of Gachon University Gil Medical Center (IRB Number: GCIRB2020-477) approved this study and the requirement for informed consent was waived because of the retrospective nature of the study design. All experimental protocols were performed in accordance with the relevant guidelines and regulations in compliance with the Declaration of Helsinki.

Additional information

Correspondence and requests for materials should be addressed to K.G.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025