Computerized Segmentation Method for Nonmasses on Breast DCE-MRI Images Using ResUNet++ with Slice Sequence Learning and Cross-Phase Convolution

Akiyoshi Hizukuri[1] · Ryohei Nakayama[1] · Mariko Goto[2] · Koji Sakai[2]

Abstract

The purpose of this study was to develop a computerized segmentation method for nonmasses using ResUNet++ with a slice sequence learning and cross-phase convolution to analyze temporal information in breast dynamic contrast material-enhanced magnetic resonance imaging (DCE-MRI) images. The dataset consisted of a series of DCE-MRI examinations from 54 patients, each containing three-phase images, which included one image that was acquired before contrast injection and two images that were acquired after contrast injection. In the proposed method, the region of interest (ROI) slice images are first extracted from each phase image. The slice images at the same position in each ROI are stacked to generate a three-dimensional (3D) tensor. A cross-phase convolution generates feature maps with the 3D tensor to incorporate the temporal information. Subsequently, the feature maps are used as the input layers for ResUNet++. New feature maps are extracted from the input data using the ResUNet++ encoders, following which the nonmass regions are segmented by a decoder. A convolutional long short-term memory layer is introduced into the decoder to analyze a sequence of slice images. When using the proposed method, the average detection accuracy of nonmasses, number of false positives, Jaccard coefficient, Dice similarity coefficient, positive predictive value, and sensitivity were 90.5%, 1.91, 0.563, 0.712, 0.714, and 0.727, respectively, larger than those obtained using 3D U-Net, V-Net, and nnFormer. The proposed method achieves high detection and shape accuracies and will be useful in differential diagnoses of nonmasses.

Keywords Nonmass · Slice sequence learning · Cross-phase convolution · Convolutional neural network · Breast magnetic resonance imaging

Introduction

Breast cancer is the most commonly diagnosed cancer among women. In 2020, approximately 684,000 women died of breast cancer globally [1]. The early detection and treatment of breast cancer are critical. For example, Reynolds et al. [2] reported that 95% of patients were completely cured when breast cancer was detected and treated early.

✉ Akiyoshi Hizukuri
   hizukuri@fc.ritsumei.ac.jp

1 Department of Electronic and Computer Engineering, Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan

2 Department of Radiology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, 465 Kajiicho, Kawaramachi Hirokoji, Kamigyoku, Kyoto 602-8566, Japan

A nonmass where a mass has not yet formed is an important indicator for breast cancers in breast dynamic contrast material-enhanced magnetic resonance imaging (DCE-MRI) images. However, distinguishing whether a nonmass lesion is malignant or benign is difficult for radiologists [3]. For example, Baltzer et al. [4] reported that the positive predictive value (PPV) of nonmasses in DCE-MRI was quite low. Unnecessary biopsies can also cause physical problems and financial burdens on patients.

Researchers have developed computer-aided diagnosis (CADx) schemes to distinguish between benign and malignant breast nonmasses to improve the PPV of nonmasses in breast DCE-MRI. Newell et al. [5] developed a CADx scheme based on an artificial neural network with morphological, textural, and kinetic features to distinguish between benign and malignant nonmasses on DCE-MRI. Tan et al. [6] and Ayatollahi et al. [7] also used machine learning techniques with texture features on DCE-MRI images to

distinguish between benign and malignant nonmasses. Li et al. [8] and Zhou et al. [9] developed computerized classification methods for benign and malignant nonmasses using radiomic features. In these methods, all features are extracted from segmented nonmass regions. Thus, it is necessary to segment the nonmasses on DCE-MRI images to evaluate the likelihood of malignancy of the nonmasses.

Cancer patterns tend to show rapid early enhancement (wash-in), followed by a loss of enhancement (wash-out) on DCE-MRI images over time [10]. Nonmasses exist in the slice images and through-plane direction in DCE-MRI images. Therefore, to segment nonmass regions accurately, it is necessary to use a computerized method to analyze the dynamic changes in the signal intensity and the relationship between consecutive slices in both lesions. The main objective of this study was to develop a computerized segmentation method for nonmasses in breast DCE-MRI images using ResUNet++ and a combination of slice sequence learning to analyze the sequential information of consecutive slices and cross-phase convolution to incorporate the dynamic changes in the lesion signal intensity. The main contributions of this study can be summarized as follows:

(1) We propose a computerized segmentation method for nonmasses using ResUNet++ [11, 12] using cross-phase convolution to analyze the temporal information among DCE-MRI images acquired at different times and slice sequence learning to examine the sequential information between continuous slices.

(2) We show that cross-phase convolution can analyze the temporal information, whereas slice sequence learning can analyze the sequential information between continuous slices.

(3) We demonstrate that the segmentation accuracies are improved using the proposed network, ResUNet++ with the cross-phase convolution and slice sequence learning, compared with those obtained by the original ResUNet++, ResUNet++ with cross-phase convolution, ResUNet++ with slice sequence learning, 3D U-Net [13], V-Net [14], and nnFormer [15].

The remainder of this paper is organized as follows. The "Related Work" section presents an overview of related studies on the segmentation task of masses/nonmasses on DCE-MRI images. Our dataset is outlined in the "Materials" section, and a detailed explanation of the proposed method is presented in the "Methods" section. The results are described in detail in the "Experimental Results" section. The "Discussion" section provides a comparative analysis with previous methods. Finally, the conclusion and limitations of this paper are described in the "Conclusions" section.

Related Work

As mass lesions in DCE-MRI generally have clear boundaries, they can be detected and segmented in a relatively straightforward manner. However, the specificity of masses in DCE-MRI is low, typically ranging from 30 to 70% [16, 17]. In contrast, nonmass lesions exhibit a heterogeneous appearance in DCE-MRI because the tumorous tissues and stroma are mixed. In addition, the boundaries of nonmasses are generally indistinct. Therefore, the detection and segmentation of nonmasses is extremely challenging [18]. The PPV of nonmasses has also been reported to be lower than that of masses [4, 19].

Several researchers have attempted to develop CADx schemes to distinguish between benign and malignant breast lesions [5–9]. As the first step, segmentation methods for breast lesions in DCE-MRI images have been established [9, 20–29]. These methods are primarily divided into image-processing-based and deep-learning-based methods, including convolutional neural networks (CNNs).

In terms of image-processing-based methods, Zhou et al. [9] used a fuzzy C-means clustering method to segment mass regions. Shokouhi et al. [20] also proposed a segmentation method for masses in DCE-MRI using region growing based on the fuzzy C-means clustering method, which enables each pixel to belong to multiple classes with varying degrees of membership. However, in the aforementioned studies, regions of interest (ROIs) containing masses were required in advance. Zheng et al. [21] proposed a graph cut–based method for mass segmentation. This method segments mass regions by minimizing the energy function related to the similarity and segmentation smoothness of the pixels. One limitation of this method is that it requires manual initialization to provide the seeds or ROI for the foreground and background.

In terms of deep learning–based methods, Carvalho et al. used SegNet [22] and U-Net [23] for mass segmentation [24]. Dalmış et al. [25] proposed a computerized segmentation method for breast and fibroglandular tissue in DCE-MRI using two consecutive U-Nets. This method first segments the breast in the entire DCE-MRI image, which is followed by segmentation of the fibroglandular tissue inside the segmented breast. Haq et al. [26] employed conditional generative adversarial networks for mass segmentation. However, one critical limitation of these studies is that two-dimensional (2D) axial slice images were only used as inputs for the networks in the DCE-MRI. Thus, these methods cannot analyze the sequential information between continuous slices. On the other hand, Khaled et al. [13] developed an automated mass segmentation method using a 3D U-Net to analyze the axial direction information. The 3D U-Net was extended by replacing 2D operations with 3D equivalents. In this method, the

DCE-MRI volumes are divided into small patches of certain sizes for the 3D U-Net. Other researchers have also proposed 3D CNN-based segmentation methods [14, 27, 28]. Recently, some researchers have developed transformer-based networks for the segmentation task. The transformer can capture the global interactions between contexts. Qin et al. [29] proposed a two-stage breast mass segmentation model. In this method, the rough outline of the breast region is first segmented by U-Net. Based on the segmented rough outline of the breast region, a TR-IMUnet model is employed for accurate segmentation of the shape of masses. This model is based on U-Net. A transformer module, an improved dynamic rectified linear unit module, and a multi-scale parallel convolution fusion module are newly employed. Moreover, Zhou et al. [15] developed nnFormer using a 3D transformer for volumetric medical image segmentation. In this model, the local and global volume–based self-attention mechanism is newly introduced to the nnFormer for learning volume representation. The authors showed that the segmentation performance of the nnFormer was improved compared to those of previous segmentation models. However, these methods have numerous trainable parameters compared to 2D CNN, which is disadvantageous for smaller datasets. Most of the aforementioned studies focused on the segmentation of masses. To the best of our knowledge, few studies exist on segmentation methods for nonmasses [9]. In [9], radiologists determined the location and slice range of the nonmasses in breast DCE-MRI images. It would be tedious for clinicians to determine them manually in clinical practice.

Materials

DCE-MRI images from February 2010 to July 2022 were acquired using a 3-T MRI scanner (Magnetom Skyra; Siemens Healthcare) with a dedicated 16-channel breast coil at the University Hospital, Kyoto Prefectural University of Medicine (Kyoto, Japan). The inclusion criteria for this study were as follows: patients who were diagnosed as having non-mass enhancement by a board-certified radiologist with 15 years of experience in breast MRI and histopathologically confirmed to be malignant via biopsy or benign via biopsy or follow-up (at least 2 years). A total of 59 consecutive patients were identified. The exclusion criteria were patients receiving any prior treatment for breast cancer ($n = 2$), insufficient image quality ($n = 2$), and breast lymphoma ($n = 1$). Finally, the database consisted of 54 DCE-MRI examinations, which contained three sequential phase images from 54 patients (mean age: 55.2 years, age range: 21–85 years).

A 3D MRI was obtained as a DCE-MRI before and two times after bolus injection of a contrast agent. Two post-contrast scans were performed with the k-space centered at 90 s (early phase) and 300 s (delayed phase) following contrast injection. The one pre-contrast and two post-contrast series generated images with a spatial resolution of $0.91 \times 0.91 \times 1.0$ mm^3 and a data matrix of 352×352 pixels. Each of the three image scan series consisted of 144 slices. A total of 63 nonmasses were included. A board-certified radiologist with 15 years of experience in breast MRI manually determined the nonmass mask images using the 3D Slicer software (https://www.slicer.org). Figure 1 shows examples of a pre-contrast DCE-MRI image, post-contrast DCE-MRI images of the early and delayed phases, and mask image.

A k-fold cross-validation method [30] with $k = 5$ and a patient-level split was used to train and test the proposed method. The dataset of 54 patients was randomly divided into five groups on the patient level, and the number of patients in each group was approximately equal. One group was used for the test dataset, whereas the remaining four were used for the training dataset. This process was repeated five times until each group was formed as a test dataset.
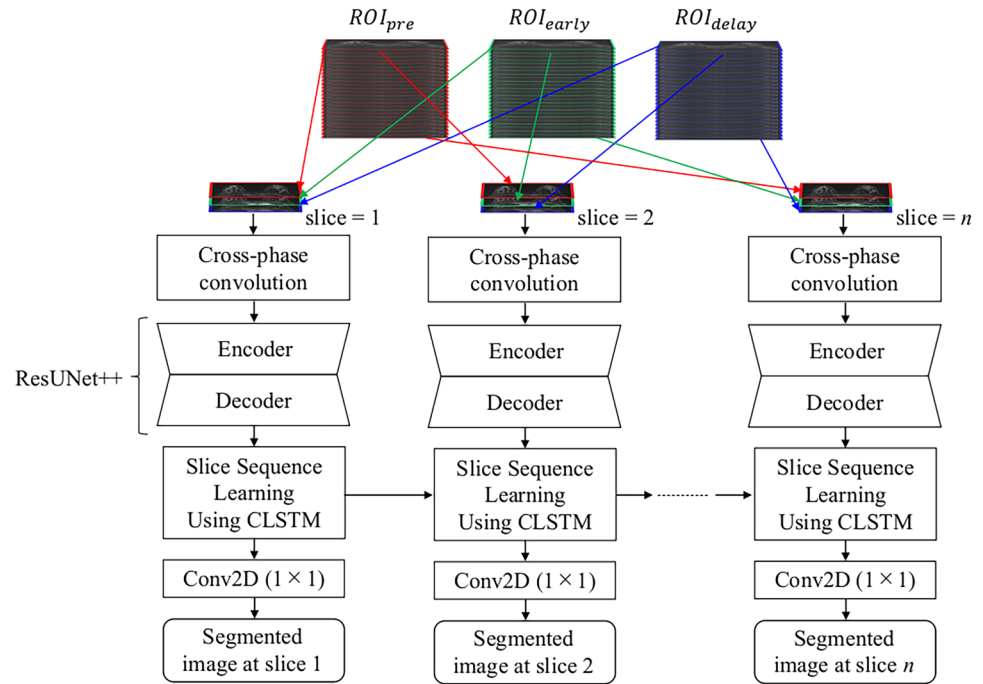
Methods

Figure 2 presents an overview of the proposed network, which primarily consists of three parts: ResUNet++ [11, 12], a cross-phase convolution to analyze dynamic changes in the lesion signal intensity, and slice sequence learning to analyze the sequential information for consecutive slice images in DCE-MRI. The details of the proposed network are presented in the following section.

Fig. 1 Example of pre-contrast DCE-MRI image, post-contrast DCE-MRI images, and mask image

Before injection of a contrast agent

Post-contrast DCE-MRI image of early phase

Post-contrast DCE-MRI image of delayed phase

Mask image
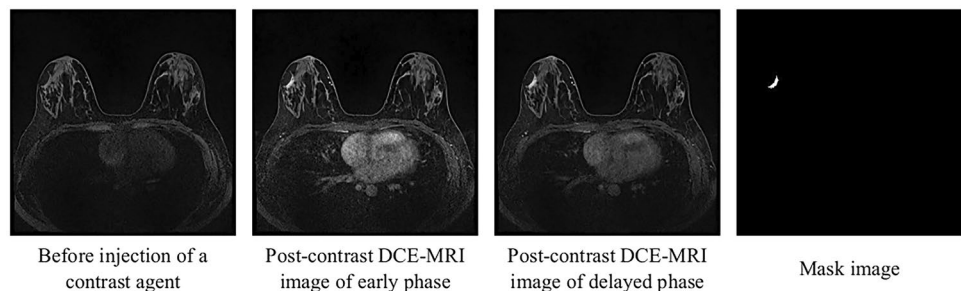
Fig. 2 Overview of the proposed network
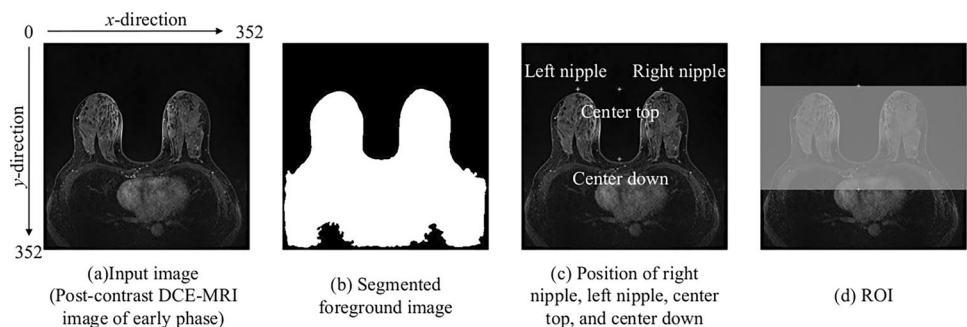

Data Augmentation

A CNN requires sufficient training data to achieve high segmentation accuracy. However, the number of training data in the database was limited. As a small amount of training data may cause the CNN to overfit; in this study, the amount of training data was doubled using horizontal flipping [31]. This data augmentation approach was employed only for the training datasets.

Extraction of Breast Region

The breast regions were extracted from the pre-contrast and two post-contrast DCE-MRI images to reduce the effects on other structures, such as the chest. The foreground, including the breast region, was segmented by applying a gray-level thresholding technique [31] to the post-contrast

DCE-MRI early-phase images. The threshold was empirically set to a 10-pixel value. Figures 3a and b depict examples of the input and segmented foreground images, respectively. Subsequently, the position (py) in the y-direction for the nipple regions was determined using the smaller value of the left nipple position in the y-direction (py_{left}) and the right nipple position in the y-direction (py_{right}). The search range was 0–$W/2$ in the x-direction, 0–$H/2$ in the y-direction, and 0 for the number of slices (S) in the z-direction (through-plane), where W, H, and S represent 352, 352, and 144 slice images, respectively. Conversely, the search range for the x-direction was set to $W/2$–W to calculate the right nipple position (px_{right}, py_{right}). The search ranges for the y- and z-directions were identical to those used to calculate the left nipple position. Subsequently, the top-center position (cx_{top}, cy_{top}) between the left and right nipples was calculated using the following equation:

Fig. 3 Example of foreground segmentation

(a) Input image (Post-contrast DCE-MRI image of early phase)

(b) Segmented foreground image

(c) Position of right nipple, left nipple, center top, and center down

(d) ROI

$$\left(cx_{top}, cy_{top} \right) = \left(\frac{nlx_{top} + nrx_{top}}{2}, py \right) \qquad (1)$$

The bottom center position (cx_{bottom}, cy_{bottom}) of the breastbone with the minimum value in the y-direction was determined based on the top center position according to the segmented breast region using the raster scan technique. The search range was cx_{top} in the x-direction, cy_{top} –H in the y-direction, and 0–S in the z-direction. Figure 3c shows an example of the left nipple, right nipple, top center, and bottom center positions. The cropping area position for the breast region was determined in the range of 1–W in the x-direction, (cy_{top}-$interval_{top}$)–(cy_{bottom}+$interval_{down}$) in the y-direction, and 1–S in the z-direction. In this case, $interval_{top}$ and $interval_{down}$ were empirically set to 5 and 50 pixels, respectively. Based on the position, ROI_{pre}, ROI_{early}, and ROI_{post} including the breast region were extracted from the pre- and post-contrast DCE-MRI images of the early and delayed phases, respectively. Figure 3d shows the clipping area range in light white. Each ROI was resized to 192×352 pixels.

Cross-Phase Convolution

The cross-phase convolution, which consisted of a 3D convolutional layer, rectified linear unit (ReLU) function, and batch normalization layer, was developed to obtain the optimal fusion among ROIs that were acquired at different times through the network training. Slice images of the same position from the ROIs at the pre-, early-, and delayed phases were first stacked together to generate a 3D tensor. The size of the 3D tensor was three phases $\times 192 \times 352$ pixels. The 3D convolution layer, ReLU function, and batch normalization layer were sequentially applied to the 3D tensor. The 3D convolution layer kernel size was $3 \times 1 \times 1$, where 3 is the number of phases. The number of filters in the 3D convolution layer was 16. This 3D layer was designed to assign weights to each ROI by training and summing their values. New feature maps were generated based on optimal fusion among the ROIs acquired at different times by cross-phase convolution.

ResUNet++

Our database, which included 54 patient examinations, was relatively small. Jha et al. [12] showed that ResUNet++ worked well with a smaller number of images. In their experiments, ResUNet++ also outperformed the well-known segmentation architectures U-Net and ResUNet. Therefore, ResUNet++ was used as the baseline network to segment nonmasses in DCE-MRI images.

ResUNet++ contained four important components: a residual block, a squeeze and excitation (SE) block, atrous spatial pyramid pooling (ASPP), and an attention block. Figure 4
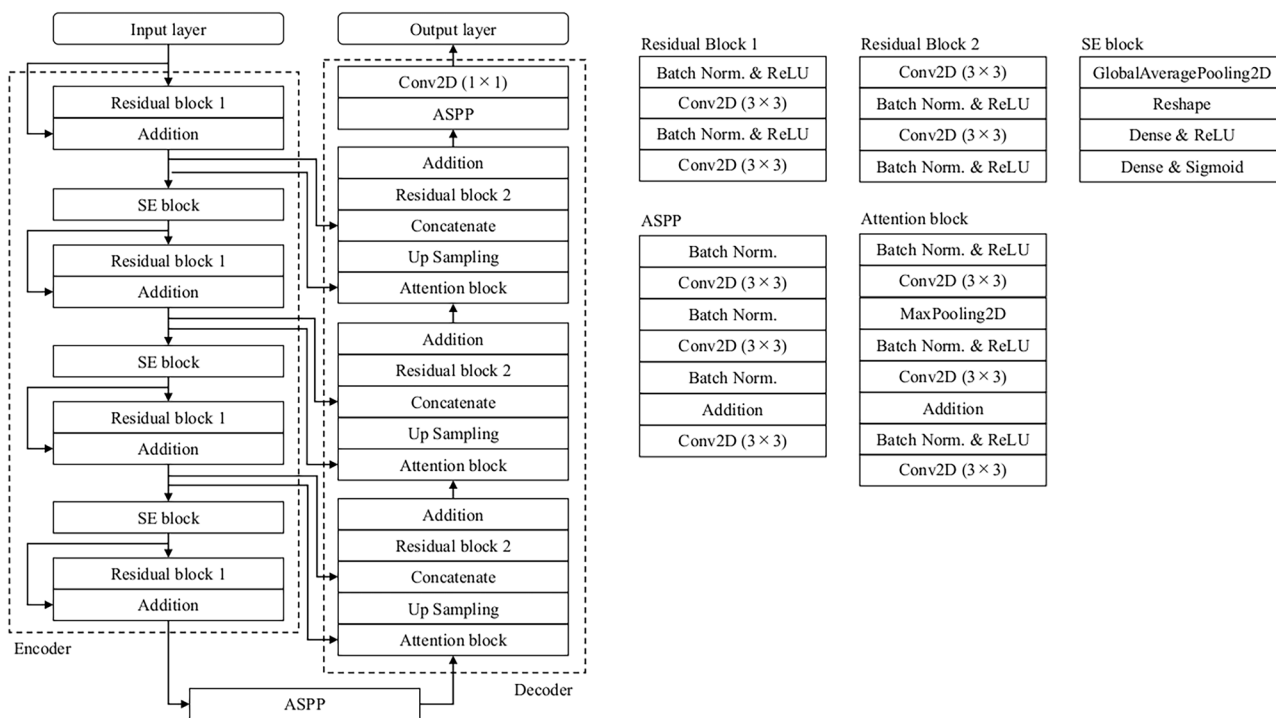

Fig. 4 Architecture of ResUNet++

shows the ResUNet++ architecture employed in this study. This network had an input layer, four encoder blocks, an SE block, ASPP, three decoder blocks, and an output layer. The feature maps that were generated by the cross-phase convolution were first input into ResUNet++. Each residual block consisted of two batch normalization layers, a ReLU function, and convolutional layers. Note that only the first residual block consisted of two convolutional layers: a batch normalization layer and a ReLU function. Skip connections connected the input and output of the residual blocks to prevent the vanishing gradient problem. The residual block outputs, excluding the final residual block, were fed into the SE blocks. The SE block learned the importance of different feature channels and adaptively recalibrated them. The feature maps obtained by the encoders were passed through the ASPP, which acted as a bridge between the encoder and decoder and could effectively aggregate contextual information at different scales without increasing the computational cost. The ASPP output was input into the decoder. The decoder consisted of attention and residual blocks and the ASPP. The attention block, which determined which parts of the images the network focused on, was executed before each residual block. The generated feature maps that were obtained by each attention block were upsampled by the nearest neighbor and concatenated with feature maps from their corresponding encoding path. The decoder output was fed to the ASPP. Subsequently, the nonmasses were segmented by applying a 1×1 convolutional layer to the 64-component feature vector that was obtained by the final ASPP.

Slice Sequence Learning

Convolutional long short-term memory (CLSTM) [32, 33] was introduced after the decoder in ResUNet++ to analyze the sequential information between consecutive slices. Compared to traditional LSTM, CLSTM replaces matrix multiplication with a convolutional operator to preserve long-term spatial information. The feature maps that were obtained from the second convolutional layer in the decoder were used as inputs for the CLSTM. The CLSTM consisted of an input gate i_t, forget gate f_t, memory cell C_t, output gate o_t, and hidden state h_t, as follows:

$$i_t = \sigma\left(W_{xi} * x_t + W_{hi} * h_{t-1} + b_i\right) \tag{1}$$

$$f_t = \sigma\left(W_{xf} * x_t + W_{hf} * h_{t-1} + b_f\right) \tag{2}$$

$$C_t = f_t \circ C_{t-1} + i_t \tanh\left(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c\right) \tag{3}$$

$$o_t = \sigma\left(W_{xo} * x_t + W_{ho} * h_{t-1} + b_o\right) \tag{4}$$

$$h_t = o_t \circ \tanh\left(C_t\right) \tag{5}$$

where x_t, h_t, and C_t are the input, hidden state, and memory cell tensors, respectively, at time step t. In this study, the time steps represented the DCE-MRI slice images. b_i, b_f, b_o, and b_c are biased terms. W_{x*} and W_{h*} are the convolutional kernels for the input and hidden states, respectively. The CLSTM with feature maps obtained by the second-last convolutional layer in the decoder was employed to learn the consecutive slice sequential information. Subsequently, a convolutional layer with a kernel of size 1×1 was employed to map the feature maps that were obtained from the CLSTM to a binary output image (nonmass region: 1, other: 0).

Loss Function for Training Proposed Network

The proposed network loss function is shown in Eq. (6):

$$\mathcal{L} = TL_{ResUNet++} + TL_{CLSTM} \tag{6}$$

where $TL_{ResUNet++}$ is defined as the Tversky loss [34] between the output images by ResUNet++ and the mask images and TL_{CLSTM} is defined as the Tversky loss between the image output by the CLSTM and the mask images. The Tversky loss function [34] is expressed as

$$TL = 1 - \frac{\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i g_i + \alpha \sum_{i=1}^{N} \left(1 - g_i\right) p_i + \beta \sum_{i=1}^{N} g_i \left(1 - p_i\right)} \tag{7}$$

where p_i and g_i denote the nonmass regions segmented by the proposed network and the mask images at pixel i, respectively; N is the number of image pixels; and α and β are hyperparameters that control the tradeoff between false positives and negatives.

Comparison with Other Segmentation Networks

We compared the proposed network with a network using a space–time memory (STM) [35] and the cross-phase convolution (network$_{STM\_CPC}$). The STM calculates the spatio-temporal attention on every pixel in multiple slice images of DCE-MRI [35]. Here, the baseline network of the network$_{STM\_CPC}$ was the same as the proposed network.

The proposed network was also compared with 3D-based CNN models, namely 3D U-Net, V-Net, and nnFormer. In 3D U-Net and V-Net, ROI_{pre}, ROI_{early}, and ROI_{delay} were first divided into $64 \times 64 \times 64$ patches (small regions), and the mask images were divided identically at the corresponding positions. Patches in ROI_{pre}, ROI_{early}, and ROI_{delay} were used as the input layer in each network for training. The patches obtained from the mask images were used as the desired output values in the network output layer. Here, in nnFormer, ROI_{pre},

ROI_{early}, and ROI_{delay} were divided into $96 \times 96 \times 96$ patches (small regions), and the mask images were divided identically at the corresponding positions.

Proposed Network Training and Testing

The proposed network was developed and evaluated using PyTorch 1.10.0 on a workstation (CPU: Intel Core i9-9900X processor, RAM: 128 GB, and GPU: NVIDIA GeForce RTX 2080 Ti). Adam was employed to minimize the loss between the output values of the proposed network and mask images. In this case, β_1 and β_2 in Adam were 0.9 and 0.999, respectively. The hyperparameters for training the proposed network were set to an epoch number of 20, an initial learning rate of 1×10^{-4}, and a mini-batch size of 5. The α and β values in the loss function were set to 0.3 and 0.7, respectively. The same parameter values were used for network$_{STM\_CPC}$, 3D U-Net, V-Net, and nnFormer.

Evaluation of Detection and Shape Accuracy

The detection and shape accuracies of the proposed network were evaluated using the ensemble average from the test datasets over the five cross-validation methods. When the gravity of a true nonmass region determined by a radiologist was within the segmented candidate for nonmasses by the proposed network, this candidate was considered to be "truly" detected. In contrast, when a true nonmass region was not within a segmented candidate, the candidate was considered to be a false positive. The Jaccard coefficient (JC), PPV, sensitivity, and Dice similarity coefficient (DSC) were used to evaluate the shape accuracy of the segmented nonmass regions using the proposed network. These evaluation criteria are defined as follows:

$$\text{JC} = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

$$PPV = \frac{|A \cap B|}{|A|} \tag{9}$$

$$Sensitivity = \frac{|A \cap B|}{|B|} \tag{10}$$

$$DSC = \frac{2|A \cap B|}{|A| + |B|} = \frac{2 \cdot PPV \cdot Sensitivity}{PPV + Sensitivity} \tag{11}$$

where A represents the nonmass regions segmented by the proposed network, and B represents the mask images. The DSC, which is also known as the F1 score, also evaluates the PPV harmonic mean and sensitivity [36].

Experimental Results

Ablation Study

Ablation studies were conducted to investigate the effectiveness of the cross-phase convolution and slice sequence learning in the proposed network. The experimental results are listed in Table 1. The sensitivity of ResUNet++ with cross-phase convolution (0.781) was slightly lower than that of the original ResUNet++ (0.797). However, the JC, PPV, and DSC, which indicate the harmonic mean of the PPV and sensitivity, of ResUNet++ with cross-phase convolution were improved compared to those of the original ResUNet++. The detection accuracy was identical for the original ResUNet++ and ResUNet++ with cross-phase convolution. We adopted slice sequence learning for ResUNet++ to obtain the feature representations between continuous slices. Although the sensitivity of ResUNet++ with slice sequence learning (0.757) was lower than that of the original ResUNet++ (0.797); it achieved a detection accuracy of 1.59%, DSC of 3.28%, and PPV of 7.23%. The number of false positives in ResUNet++ with cross-phase convolution (3.18) and ResUNet++ with slice sequence learning (2.22) was higher than that of the original ResUNet++ (2.17). Finally, the sensitivity of ResUNet++ with both cross-phase convolution and slice sequence learning (0.727, proposed network) was lower than that of the original ResUNet++ (0.797), that with cross-phase convolution (0.781), and that with slice sequence learning (0.757). However, the remaining evaluation indices of the proposed network improved substantially, as presented in Table 1.

Figure 5 compares the nonmass region images segmented by the original ResUNet++, ResUNet++ with cross-phase convolution, ResUNet++ with slice sequence learning, and the proposed network. The segmented regions for the

Table 1 Ablation study results

| Backbone	Cross-phase convolution	Slice sequence learning	Detection accuracy	Number of false positive per patient	JC	DSC	PPV	Sensitivity
ResUNet++	-	-	87.3% (55/63)	2.17	0.479	0.647	0.545	0.797
	✓		87.3% (55/63)	3.19	0.515	0.692	0.620	0.781
		✓	88.9% (56/63)	2.22	0.529	0.680	0.617	0.757
	✓	✓	**90.5% (57/63)**	**1.91**	**0.563**	**0.712**	**0.714**	**0.727**
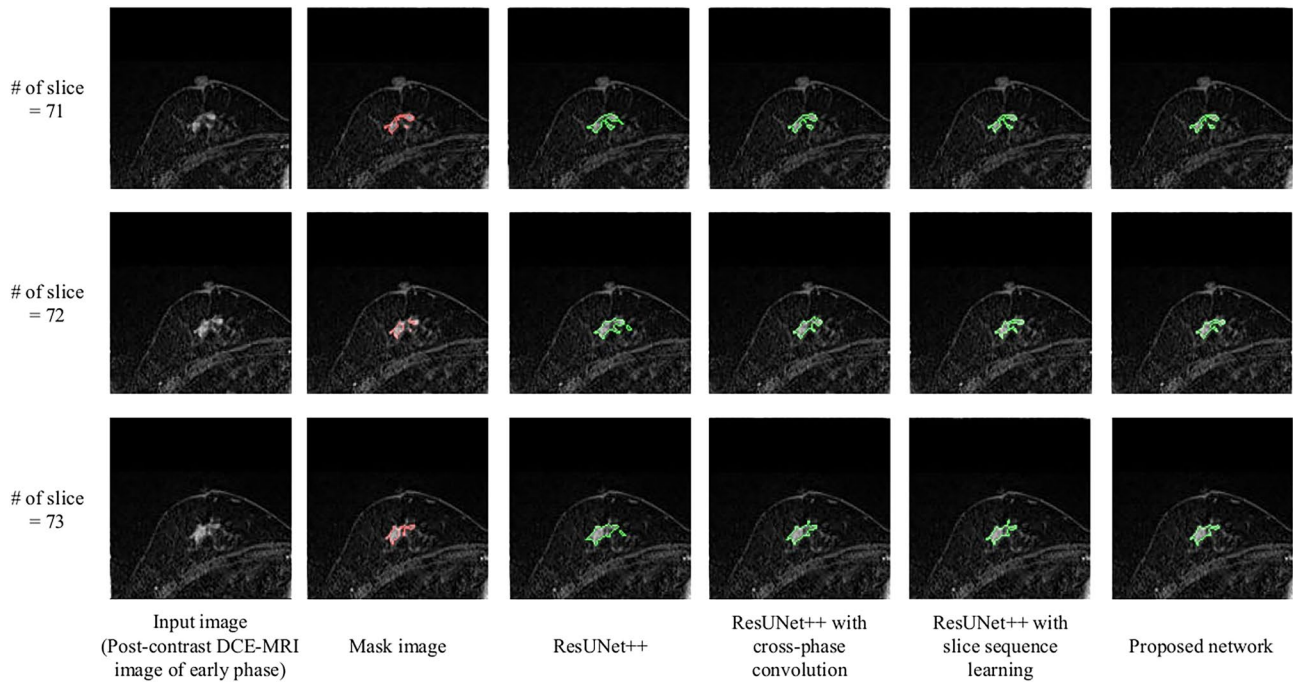
Fig. 5 Examples of nonmasses segmented by ResUNet++, ResUNet++ with cross-phase convolution, ResUNet++ with slice sequence learning, and the proposed network

original ResUNet++, ResUNet++ with cross-phase convolution, and ResUNet++ with slice sequence learning included parts of the normal tissue that were misclassified as nonmasses. However, the proposed network correctly segmented nonmasses compared to the other networks.

Comparison Results of Conventional Segmentation Networks

Figure 6 shows an example of segmented nonmass regions using the proposed network, network$_{STM\_CPC}$, 3D U-Net,
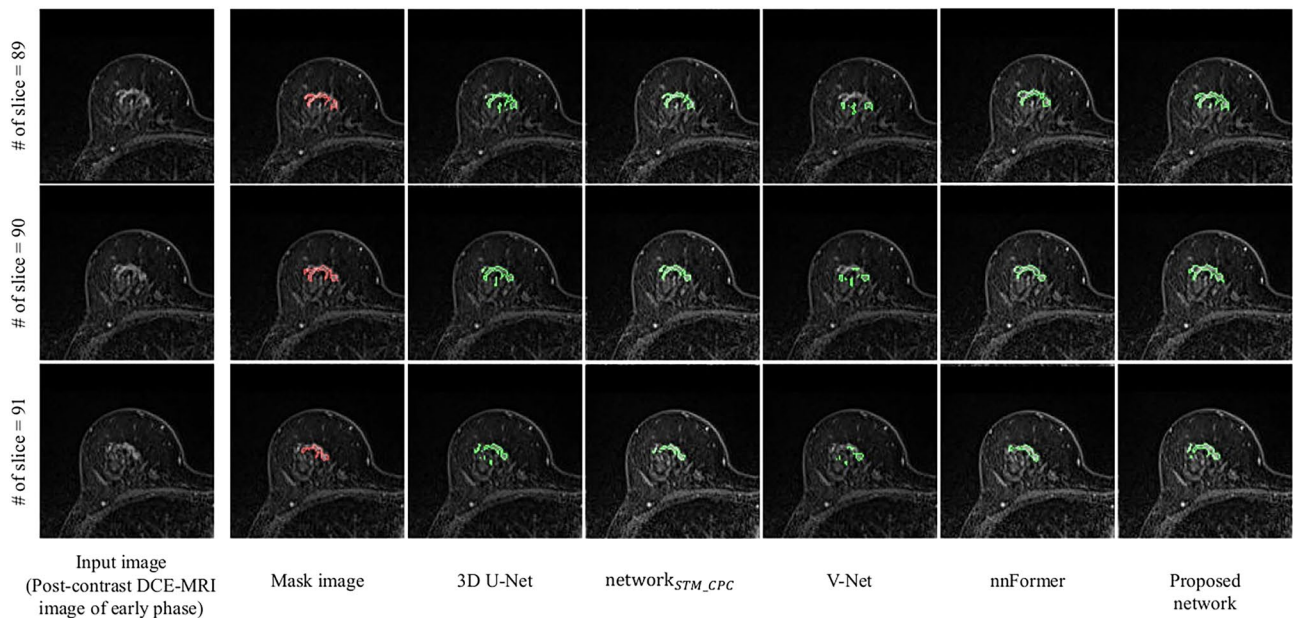
Fig. 6 Examples of segmented nonmasses by network$_{STM\_CPC}$, 3D U-Net, V-Net, nnFormer, and the proposed network

Table 2 Comparison of results obtained by network$_{STM\_CPC}$, 3D U-Net, V-Net, nnFormer, and the proposed network

Method	Detection accuracy	Number of false positive per patient	JC	DSC	PPV	Sensitivity
network$_{STM\_CPC}$	88.9% (56/63)	4.81	0.468	0.610	0.668	0.707
3D U-Net	82.5% (52/63)	1.93	0.463	0.654	0.694	0.720
V-Net	90.5% (57/63)	3.76	0.479	0.661	0.668	0.742
nnFormer	85.7% (54/63)	2.13	0.489	0.656	0.669	0.652
Proposed network	**90.5% (57/63)**	**1.91**	**0.563**	**0.712**	**0.714**	**0.727**

V-Net, and nnFormer. Table 2 shows the comparison results for network$_{STM\_CPC}$, 3D U-Net, V-Net, nnFormer, and the proposed network.

The evaluation indices of the proposed network (90.5% detection accuracy, 1.91 false positives, 0.563 JC, 0.712 DSC, 0.714 PPV, and 0.727 sensitivity) were improved compared to network$_{STM\_CPC}$ (88.9%, 4.81, 0.468, 0.610, 0.668, and 0.707). In network$_{STM\_CPC}$, the number of training images may have been relatively small for the STM because, in [35], many of the training data compared to this study were used to train the STM. Therefore, the segmentation accuracy of network$_{STM\_CPC}$ may be improved by using more training data. However, collecting a large number of DCE-MRI images containing nonmasses is generally difficult.

In the comparison of the proposed network with 3D-based CNN models, the sensitivity of the proposed network (0.727) was lower than that of V-Net (0.742). In contrast, the remaining evaluation indices of the proposed network (90.5% detection accuracy, 1.91 false positives, 0.563 JC, 0.712 DSC, and 0.714 PPV) were higher than those of 3D U-Net (82.5%, 1.93, 0.463, 0.654, and 0.694, respectively), V-Net (90.5%, 3.76, 0.479, 0.661, and 0.668, respectively), and nnFormer (85.7%, 2.13, 0.489, 0.656, and 0.669, respectively). These 3D-based CNN models cannot enhance the regions for temporal enhancement changes in the DCE-MRI images. Therefore, we believe that the proposed network is more appropriate for segmenting non-masses on breast MRI images.

Discussion

In this study, we developed a method to improve the segmentation performance of nonmasses in DCE-MRI images. Cross-phase convolution is used to analyze the temporal information among DCE-MRI images acquired at different times, and slice sequence learning is utilized to examine the sequential information between continuous slices. Segmented images of nonmass regions can be generated with higher accuracy than those obtained using conventional methods by employing the proposed method.

According to Table 1 and 2, the detection accuracy, JC, and DSC of the original ResUNet++ were lower than those

of V-Net. Nonmasses existed in the slice images as well as in the through-plane direction in the DCE-MRI. Therefore, the original ResUNet could not capture the volumetric information of nonmass lesions.

We compared ResUNet++ with the cross-phase convolution to the original ResUNet++ to investigate the benefits of the cross-phase convolution. As shown in Table 1, the sensitivity of ResUNet++ with cross-phase convolution was slightly lower than that of the original ResUNet++, whereas the PPV of ResUNet++ with cross-phase convolution was higher than that of the original ResUNet++. The DSC, which evaluated the balance between the sensitivity and PPV of ResUNet++ with cross-phase convolution, was higher than that of the original ResUNet++. The remaining evaluation indices for ResUNet++ with cross-phase convolution were also improved compared with those of the original ResUNet++. The network must learn the temporal enhancement changes of the nonmass regions to segment nonmass regions accurately. The original ResUNet++ could not be trained to focus on temporal enhancement changes in the nonmasses. By introducing cross-phase convolution to the original ResUNet++, the network could be trained by focusing on mass-enhancement changes that appeared in the DCE-MRI. Figure 7 shows the feature map with the highest mean value in the nonmass region among the feature maps that were obtained from the cross-phase convolution. It can be observed that the cross-phase convolution enhanced the regions for temporal enhancement changes in the DCE-MRI images. Some studies have used the difference images obtained by subtracting the pre-contrast DCE-MRI images from the post-contrast DCE-MRI images as the network input to reflect the mass enhancement changes in the network [13]. However, the different images enhanced minute differences, including noise, between the DCE-MRI images that were acquired at different times. Therefore, the use of different images may have a negative impact on the network training.
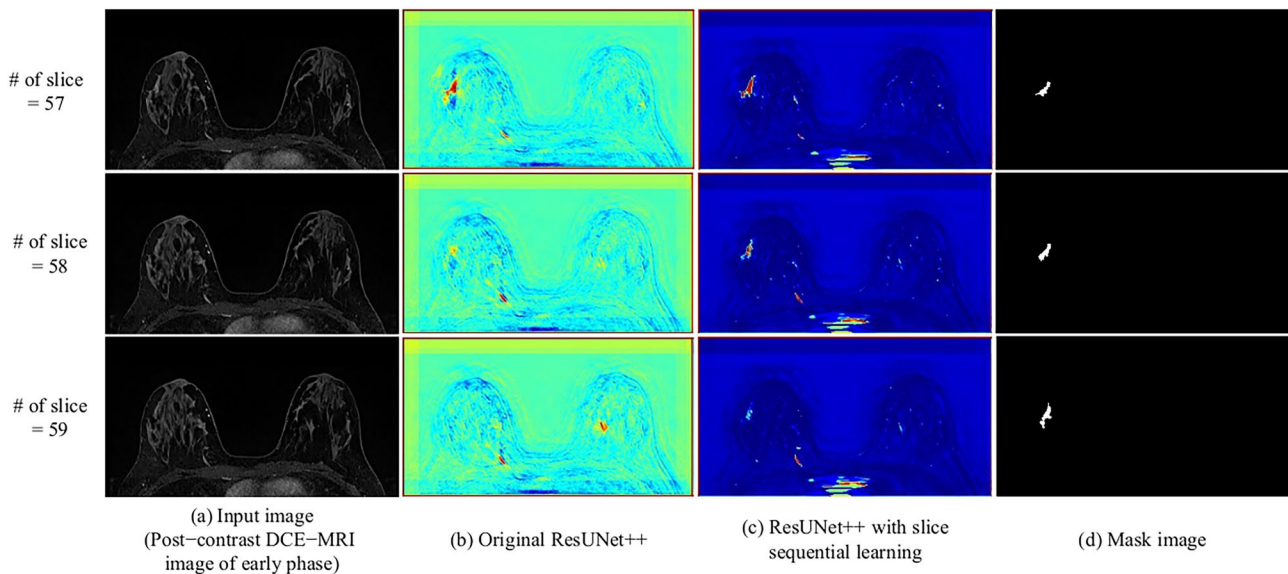
ResUNet++ with slice sequence learning was compared with the original ResUNet++ in terms of the detection and shape accuracy to investigate the benefits of slice sequence learning in analyzing the sequential information between consecutive slices. Although the sensitivity of ResUNet++ with slice sequence learning was lower than that of

Fig. 7 Visualization of feature map from cross-phase convolution

(a) ROI_{pre}

(b) ROI_{early}

(c) ROI_{delay}

(d) Feature map by cross-phase convolution
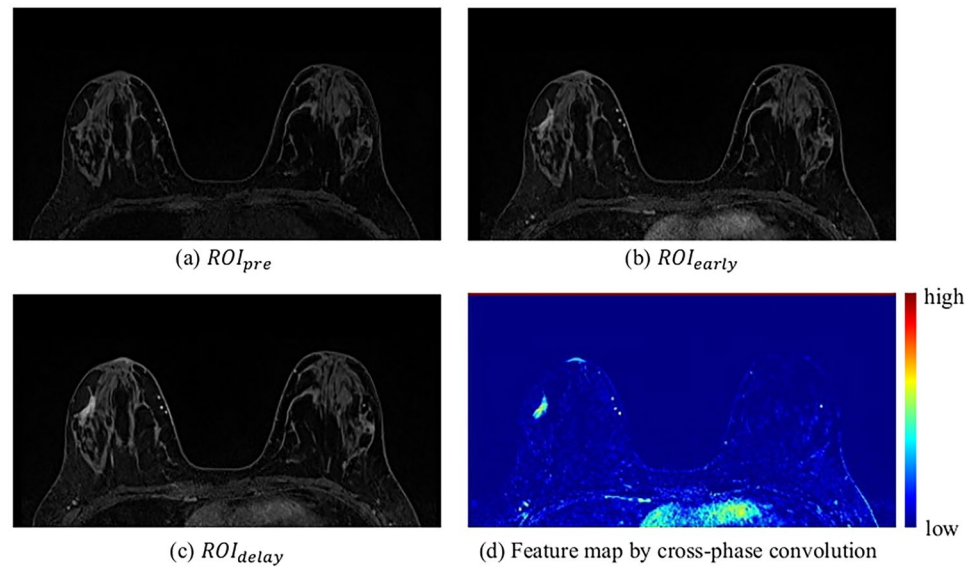
the original ResUNet++, the remaining evaluation indices for ResUNet++ with slice sequence learning were improved. Figure 8 shows the feature maps that were obtained using the CLSTM. The mean value of each feature map was calculated and visualized with the highest mean value in the nonmass region among all feature maps. The results showed that the CLSTM enhanced the features between consecutive slices that contained nonmasses compared to the original ResU-Net++. Therefore, analysis of the through-plane direction enables the network to consider 3D contextual information, which assists the network in better distinguishing between nonmasses and normal tissue, especially for nonmasses that appear similar to normal tissue when viewed in a single slice.

Some limitations of this study should be noted. One is that the JC of the proposed network was relatively low. It is well known that the detection and segmentation of non-masses are extremely challenging for radiologists because of their poorly defined boundaries compared to masses. If radiologists slightly revise the segmented nonmass images that are obtained by the proposed network, we believe that these images can be used in CADx schemes to evaluate the likelihood of malignancy of nonmasses. Therefore, when radiologists use CADx schemes, the proposed method can decrease the burden compared to manual tracing. Another limitation is that hyperparameters such as the mini-batch size, epoch number, and learning rate in the proposed network may not have been the most appropriate combination

# of slice = 57

# of slice = 58

# of slice = 59

(a) Input image (Post−contrast DCE−MRI image of early phase)

(b) Original ResUNet++

(c) ResUNet++ with slice sequential learning

(d) Mask image

Fig. 8 Visualization of feature maps from CLSTM and output of decoder in original ResUNet++

for the detection and segmentation of nonmass regions. Thus, the detection and segmentation accuracy may be improved by using more suitable hyperparameter combinations. Finally, data from only 54 patient examinations were used in this study. Therefore, future research will focus on expanding the database and evaluating the performance of the proposed network on this basis.

Conclusions

We developed a computerized segmentation method for nonmasses in breast DCE-MRI using ResUNet++ combined with slice sequence learning and cross-phase convolution. The experiment results showed that the slice sequence learning analyzes the sequential information of consecutive slices, and the cross-phase convolution can capture the dynamic changes in the lesion signal intensity.

The proposed network exhibited a higher segmentation accuracy than the original ResUNet++, ResUNet++ with cross-phase convolution, and ResUNet++ with slice sequence learning. It also outperformed 3D U-Net, V-Net, and nnFormer. Thus, the proposed network may be useful for segmenting nonmasses in breast DCE-MRI.

Some CADx schemes for evaluating the likelihood of malignancy of nonmasses require nonmass mask images. Although radiologists manually determine the mass regions using the CADx schemes, it would be tedious for them to trace masses manually in clinical practice. Thus, when radiologists utilize CADx schemes, the proposed method can save time compared to manual tracing.

Future works will include improving the segmentation accuracy of the proposed network by optimizing the hyperparameters using Bayesian optimization. We also will introduce the proposed method into CADx schemes to evaluate the likelihood of malignancy of nonmasses and whether the classification performance is improved compared to previous segmentation methods. Moreover, we will focus on expanding the database and evaluating the performance of the proposed network on this basis.

Declarations

References

1. H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: Cancer J. Clin., vol. 71, no. 3, pp. 209-249, Jun. 2021.

2. H.E. Reynolds, V.P. Jackson, "Self-referred mammography patients: analysis of patients' characteristics," Am. J. Roentgenol., vol. 157, no. 3, pp. 48-484, Jan. 1991.

3. A. Meyer-Base, L. Morra, A. Tahmassebi, M. Lobbes, U. Meyer-Base, K. Pinker, "AI-enhanced diagnosis of challenging lesions in breast MRI: A methodology and application primer," J. Magn. Reson. Imaging, vol. 54, no. 3, pp. 686-702, 2021.

4. P.A. Baltzer, M. Benndorf, M. Dietzel, M. Gajda, I.B. Runnebaum, W.A. Kaiser, "False-positive findings at contrast-enhanced breast MRI: a BI-RADS descriptor study," Am. J. Roentgenol., vol. 194, no. 6, pp. 1658-1663, Jun. 2009.

5. D. Newell, K. Nie, J.H. Chen, C.C. Hsu, H.J. Yu, O. Nalcioglu, M.Y. Su, "Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement," Eur. Radiol., vol. 20, no. 4, pp. 771-781, Apr. 2010.

6. Y. Tan, H. Mai, Z. Huang, L. Zhang, C. Li, S. Wu, K. Jiang, "Additive value of texture analysis based on breast MRI for distinguishing between benign and malignant non-mass enhancement in premenopausal women," BMC Med. Imaging, vol. 21, no. 1, pp. 1-10, Mar. 2021.

7. F. Ayatollahi, S.B. Shokouhi, J. Teuwen, "Differentiating benign and malignant mass and non-mass lesions in breast DCE-MRI

using normalized frequency-based features," Int. J. Comput. Assist. Radiol. Surg., vol. 15, no. 2, pp. 297-307. Feb. 2020.

8. Y. Li, Z.L. Yang, W.Z. Lv, Y.J. Qin, C.L. Tang, X. Yan, et.al. "Non-mass enhancements on DCE-MRI: development and validation of a radiomics-based signature for breast cancer diagnoses," Front. Oncol., vol. 11, pp. 1-12, Sep. 2021.

9. J. Zhou, Y.L. Liu, Y. Zhang, et al. "BI-RADS reading of non-mass lesions on DCE-MRI and differential diagnosis performed by radiomics and deep learning," Front. Oncol., vol. 11, pp. 1-10, Nov. 2021.

10. American College of Radiology, Breast imaging reporting and data system (BI-RADS), 5th edition, American College of Radiology, 2013.

11. D. Jha, P.H. Smedsrud, D. Johansen, T. de Lange, H.D. Johansen, P. Halvorsen, M.A. Riegler, "A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation," IEEE J. Biomed. Health. Inform., vol. 25, no. 6, pp. 2029-2040, Jun. 2021.

12. D. Jha, P.H. Smedsrud, M.A. Riegler, D. Johansen, T. de Lange, P. Halvorsen, H.D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," IEEE Int. Symp. Multimedia (ISM), pp. 225–2255, Dec. 2019.

13. R.A. Khaled, J. Vidal, R. Martí, "Deep learning based segmentation of breast lesions in DCE-MRI," Pattern Recognition. ICPR International Workshops and Challenges, Part I, pp. 417-430, Jan. 2021.

14. F. Milletari, N. Navab, S.A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," Int. Conf. 3D Vis. (3DV), pp. 565–571, Oct. 2016.

15. H.Y. Zhou, J. Guo, Y. Zhang, X. Han, L. Yu, L. Wang, Y. Yu, "nnFormer: volumetric medical image segmentation via a 3D transformer," vol.32, pp.4036–4045, IEEE Transactions on Image Processing, 2023.

16. R. Fusco, M. Sansone, S. Filice, G. Carone, D.M. Amato, C. Sansone, A. Petrillo, "Pattern recognition approaches for breast cancer DCE-MRI classification: a systematic review," J. Med. Biol. Eng., vol. 36, pp. 449-459, 2016.

17. A. Hizukuri, R. Nakayama, M. Nara, M. Suzuki, K. Namba, "Computer-aided diagnosis scheme for distinguishing between benign and malignant masses on breast DCE-MRI images using deep convolutional neural network with Bayesian optimization," J. Digit. Imaging, vol. 34, pp. 116-123, 2021.

18. H. Yabuuchi, Y. Matsuo, T. Kamitani, et al., "Non-mass-like enhancement on contrast-enhanced breast MRI imaging: Lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images," Eur. J. Radiol., vol. 75, no. 1, pp. 126-132, 2010.

19. T. Asada, T. Yamada, Y. Kanemaki, K. Fujiwara, S. Okamoto, Y. Nakajima, "Grading system to categorize breast MRI using BI-RADS 5th edition: a statistical study of non-mass enhancement descriptors in terms of probability of malignancy," Jpn. J. Radiol., vol. 36, pp. 200-208, 2018.

20. S.B. Shokouhi, A. Fooladivanda, N. Ahmadinejad, "Computer-aided detection of breast lesions in DCE-MRI using region growing based on fuzzy C-means clustering and vesselness filter," EURASIP J. Adv. Signal Process., vol. 1, pp. 1-11, May 2017.

21. Y. Zheng, S. Baloch, S. Englander, M.D. Schnall, D. Shen, "Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images," Med. Image Comput. Comput. Assist. Interv., vol. 10 (part II), pp. 393–401, Oct. 2007.

22. V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481-2495, Dec. 2017.

23. O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," Med. Image Comput. Comput. Assist. Interv., vol. 18 (part III), pp. 234–241, Oct. 2015.

24. E.D. Carvalho, R.R.V. Silva, M.J. Mathew, F.H.D. Araujo, A.O. De Carvalho Filho, "Tumor segmentation in breast DCE-MRI slice using deep learning methods," IEEE Symp. Comput. Commun. (ISCC), pp.1–6, Sep. 2021.

25. M.U. Dalmış, G. Litjens, K. Holland, A. Setio, R. Mann, N. Karssemeijer, A. Gubern-Mérida, "Using deep learning to segment breast and fibroglandular tissue in MRI volumes," Med. Phys., vol. 44, no. 2, pp. 533-546, Feb. 2017.

26. I.U. Haq, H. Ali, H.Y. Wang, L. Cui, J. Feng, "BTS-GAN: computer-aided segmentation system for breast tumor using MRI and conditional adversarial networks," Eng. Sci. Technol. Int. J., vol. 36, pp. 1-10, 2022.

27. M. Qiao, C. Li, S. Suo, F. Cheng, et al., "Breast DCE-MRI radiomics: a robust computer-aided system based on reproducible BI-RADS features across the influence of datasets bias and segmentation methods," Int. J. Comput. Assist. Radiol. Surg., vol. 15, no. 6, pp. 921-930, Jun. 2020.

28. S. Wang, K. Sun, L. Wang, L. Qu, F. Yan, Q. Wang, D. Shen, "Breast tumor segmentation in DCE-MRI with tumor sensitive synthesis," IEEE Trans. Neural. Netw. Learn. Syst., Dec. 2021.

29. C. Qin, Y. Wu, J. Zeng, L. Tian, Y. Zhai, F. Li, X. Zhang, "Joint transformer and multi-scale CNN for DCE-MRI breast cancer segmentation," Soft Computing, vol. 26, no. 17, pp.8317-8334, 2022.

30. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," IJCAI, vol. 14, no. 2, pp. 1137-1145, Aug. 1995.

31. R.C. Gonzales, R.E. Woods, Digital Image Processing, 2nd edition, Addison-Wesley, MA, pp. 567-643, 1992.

32. X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, W.C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," Adv. Neural Inf. Process. Syst. (NeurIPS), 28, 2015.

33. F. Xu, H. Ma, J. Sun, R. Wu, X. Liu, Y. Kong, "LSTM multimodal UNet for brain tumor segmentation," IEEE Int. Conf. Image, Vis. Comput. (ICIVC), pp. 236–240, 2019.

34. N. Abraham, N.M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019), pp. 683–687, Apr. 2019.

35. S.W. Oh, J.Y. Lee, N. Xu, S.J. Kim, "Video object segmentation using space-time memory networks," Proc. IEEE/CVF International Conference on Computer Vision, pp. 9226–9235, 2019.

36. K.B. Soulami, N. Kaabouch, M.N. Saidi, A. Tamtaoui, "Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation," Biomed. Signal Process. Control, vol. 66, 102481, 2021.