Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(4): 3578–3597.
DOI: 10.3934/mbe.2021180
Received: 21 October 2020
Accepted: 22 February 2021
Published: 25 April 2021

Research article

Design and analysis of a robust breast cancer diagnostic system based on multimode MR images

Hong Yu[1], Wenhuan Lu[2], Qilong Sun[3], Haiqiang Shi[4,*], Jianguo Wei[2,3] Zhe Wang[5], Xiaoman Wang[2] and Naixue Xiong[2]

[1] Center Obstetrics and Gynecology Hospital, Tianjin 300100, China
[2] College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[3] School of Computer Science, Qinghai Nationalities University, Xining Qinghai, 810007, China
[4] Qinghai Provincial Party School, Xining Qinghai, 810007, China
[5] IBM China Company Limited, Beijing 100193, China

* **Correspondence:** Email: shihaiqiang@ieee.org.

Abstract: In this paper, we propose a Robust Breast Cancer Diagnostic System (RBCDS) based on multimode Magnetic Resonance (MR) images. Firstly, we design a four-mode convolutional neural network (FMS-PCNN) model to detect whether an image contains a tumor. The features of the images generated by different imaging modes are extracted and fused to form the basis of classification. This classification model utilizes both spatial pyramid pooling (SPP) and principal components analysis (PCA). SPP enables the network to process images of different sizes and avoids the loss due to image resizing. PCA can remove redundant information in the fused features of multi-sequence images. The best accuracy of this model achieves 94.6%. After that, we use our optimized U-Net (SU-Net) to segment the tumor from the entire image. The SU-Net achieves a mean dice coefficient (DC) value of 0.867. Finally, the performance of the system is analyzed to prove that this system is superior to the existing schemes.

Keywords: classification; convolutional neural networks; Magnetic Resonance imaging; multiple modes; segmentation

1. Introduction

Breast cancer is the most prevalent type of cancer among women worldwide and it also threatens men's health [1]. More than 1.7 million new cases of breast cancer are detected around the world each year [2]. Early detection of breast cancer is an important step to reduce negative effects. The main

medical methods for breast cancer detection are X-ray mammography (mammography), Computed Tomography (CT), Doppler ultrasonography, and Magnetic Resonance imaging (MRI). Mammography has a low detection rate for dense breasts [3]. The diagnostic radiation of CT is larger than it of ordinary X-ray machines. Ultrasound devices have poor penetration into bone and it is difficult for obese patients to produce effective images with high quality.

MRI is a technique for clinical imaging of the mammary glands. Breast MRI is employed for 1) screening of breast cancer in high-risk subjects, 2) evaluation of tumor extent in specific groups of breast cancer patients (e.g., patients with invasive lobular carcinoma), 3) evaluation of tumor response to chemotherapy treatment, and 4) troubleshooting in case of inconclusive findings from other modalities [4]. Breast MRI can be effective at finding unsuspected cancer in the opposite breast [5], which can facilitate early treatment of both tumors and reduce the death rate of breast cancer. Thus, we use MRI for the diagnosis of breast cancer in this paper.

Computer-Aided Diagnosis (CAD) software has been developed and used since the last century. Unfortunately, commercial CAD systems have not brought significant improvements in performance [6], especially in the last decade. The actual diagnosis is still more dependent on the experience of clinicians. There is always a difference in the diagnostic results of radiologists and CAD. Many radiologists either fail to use CAD as a second reader or only rely on the output of CAD [7]. The clinician's observation may influence the diagnostic results to some extent. Our work is to develop a system for better classification of Magnetic Resonance (MR) images and localization of tumors in MR images. As an auxiliary reader, this system can reduce radiologists' image reading time.

In previous studies, most researchers relied heavily on single-sequence images. But in the actual diagnostic process, radiologists and clinical experts observe and analyze multi-sequence images simultaneously to diagnose the patient accurately. Therefore, we use several sets of images with different imaging modes for robust breast image classification. According to the radiologists' suggestions, we chose T1W_TSE, T2W_SPARI, DWI_SSh, and DYN_eTHRIVE+C. They are written as T1W, T2W, DWI, and DYN in the following. We use the characteristics of these four sequence images to determine whether an image contains a tumor. In addition to image classification, radiologists also need to locate tumors to lay the foundation for subsequent diagnosis and treatment. Nevertheless, the poor signal-to-noise-ratio and faint edges caused by partial volume effects increase the difficulty of tumor segmentation [8]. Different radiologists may determine different tumor locations following their own experiences. To assist radiologists in locating the tumor edge better, we propose an optimized segmentation method. The classification-segmentation process becomes the basis for radiologists to analyze and interpret MR images and clinicians to implement treatment.

Deep-learning algorithms [9–11] are capable of directly learning discriminative features from the given data, and convolutional neural networks (CNNs) have been proven to be effective models for tackling visual tasks [12–14]. Deep-learning has also produced positive results in medical disease analysis [15, 16]. In the medical image classification and segmentation task, CNNs also have extraordinary performance [17] and may perform better with the expansion of data sets and the improvement of network structure. Some scholars have focused on the robustness [18, 19] and the optimization scheme [20] of models to obtain superior and steady performance on the Internet of things. In any case, technology is increasingly tied to life. We investigate CNNs and optimization methods to produce a better solution for breast cancer diagnosis.

The main contributions of our work are:

1) We design and analyze a Robust Breast Cancer Diagnostic System (RBCDS) based on multimode MR images.

2) We propose a four-mode convolutional neural network (FMS-PCNN) model to distinguish whether there is a tumor in an image. This model combines four types of MR images comprehensively to improve experimental performance and simulate the classification process of radiologists. The spatial pyramid pooling (SPP) [21] and principal components analysis (PCA) [22] are used to utilize the original size images and remove redundancy of fused features.

3) We use the improved U-Net architecture for the segmentation task [23]. The U-Net achieves excellent performance on biomedical segmentation applications [24]. The important modification is applying the sub-pixel convolution method [25] in the upsampling process.

The rest sections of this paper are conducted as follows. Section II reviews the related work of breast cancer classification, segmentation based on MRI and CNNs in image processing. Section III describes the system model and relevant definitions. Sections IV and V show the details of classification and segmentation methods used in RBCDS. Section VI exhibits the performance of our proposed model and analyzes the comparisons across different methods. And section VII summarizes the full paper and future work.

2. Related works

2.1. Breast cancer classification

The automatic image processing for cancer diagnosis has been explored as a topic of research for more than 40 years, but it is still challenging due to the complexity of the images [26]. The state of the art in classification problems [27] on vision communities, such as action [28], object [29–31], and scene [32, 33], have been significantly improved and produced remarkable results these years.

Liu et al. [34] used the cell shape and the bare nuclei as the significant features for breast cancer diagnosis and utilized a Bayesian network (BN) modeling approach to support clinicians in making diagnostic decisions. And the BN modeling method was built by the K2 learning algorithm and statistical computation methods. They reported that the combination of all features showed the best performance, with a diagnostic accuracy of 0.934. Their model could help clinicians diagnose diseases according to important characteristics, which could be used in other healthcare data analytics. Sayed et al. [35] extracted features from breast MR images and then used K-Nearest Neighbor (KNN) and Linear Discriminant Analysis (LDA) to differentiate between benign and malignant tumors. They stated that benign cases were misclassified with an average specificity reduction percentage of about 3%. They also reported that their method may become a noninvasive diagnostic tool for breast cancer. Drukler et al. [36] assessed the performance of radiologic signatures by using a support vector machine (SVM) and 10-fold cross-validation based on a dataset consisting of 64 lesions. They calculated the 95% confidence interval of the area under the receiver operating characteristic (ROC) curve (AUC) to confirm that texture enhancement played a dominant role in the diagnosis of breast cancer. In their report, the radiologic enhancement-texture signature performed best for ultrafast MRI and achieved the AUC of 0.71.

However, in previous work of classification, only a single imaging mode was mainly investigated. Few studies involved in the analysis of multiparametric images [7, 23] and the dependencies among

features in a single experiment. Kyono et al. [7] introduced a Man and Machine Mammography Oracle (MAMMO) to support radiologists in their diagnosis using mammography datasets. They stated that multi-task learning enabled CNNs to capture more features related to cancer and their model reduced the number of radiologist reading by 42.8%. The second part of MAMMO could determine which decisions output by classification networks was correct and confident. In addition, most studies adapted the resized image to the network structures, which would obviously affect the learning of image features and experimental results. We apply SPP to classify images with original sizes, avoiding loss caused by resizing them. Most importantly, images produced by four imaging modes are used in our proposed classification model to simulate actual image reading, gather features of different sequences and improve the accuracy of classification.

2.2. Segmentation based on MRI

Breast tumors behave differently across several imaging modes in quantitative analysis. Assessment of tumor size, texture and morphology strongly relies on accurate segmentation, and it has been shown difficult even for expert radiologists.

Gubern-Mrida et al. [4] developed a method to calculate the density and segment the breast and fibro-glandular tissue, using an atlas-based approach and Expectation-Maximum (EM) algorithm. They achieved an average dice similarity coefficient (DSC) value of 0.94 ± 0.04 for breast segmentation and 0.80 ± 0.13 for fibro-glandular tissue segmentation. Maicas et al. [37] introduced a new fully automated breast tumor segmentation model. The method was based on the globally optimal inference in a continuous space (GOCS) and used a shape prior computed from a semantic segmentation. The salient advantage of their approach was that it did not require precise initialization. They achieved the mean dice coefficient (DC) of 0.77 for the test set in dynamic contrast-enhanced Magnetic Resonance imaging (DCE-MRI). Zhang et al. [38] proposed a mask-guided hierarchical learning (MHL) framework by extending the fully convolutional network (FCN), which can decompose the problem into several simpler sub-problems. This framework avoided time-consuming manual annotation for tumor regions. The segmentation was based on both segmentation probability maps and landmarks defined on nipples and achieved the mean DSC of 0.72. The most important part of the segmentation task is the localization of the tumor edge. We pay more attention to the precise location of the tumor edge by optimizing the upsampling process. Furthermore, we use DC as the evaluation coefficient of segmentation method, referring to the above studies.

2.3. Convolutional neural networks in image processing

CNNs have achieved great success in large-scale image recognition and won the large scale visual recognition challenge (ILSVRC) since 2012 [10]. Due to lack of data, however, a large number of parameters and multiple trainable layers in CNNs cannot reasonably be trained.

Hence, Ronneberger et al. [24] used data augmentation and training strategy to build the U-Net structure. The network consisted of a contracting path and an expansive path, which gave it the u-shaped architecture. Their network can be trained end-to-end from very few images. They marked a very significant achievement and won the ISBI challenge for the segmentation of neuronal structures in electron microscopic stacks in 2015.

Alom et al. [39] proposed a Recurrent Convolutional Neural Network (RCNN) and a Recurrent

Residual Convolutional Neural Network (RRCNN). The new models were all based on U-Net, which had great performance on medical image detection tasks. The recurrent residual convolutional layers could improve the feature representation and parameter design for medical image segmentation. Their method performed better on lung lesion segmentation, blood vessel segmentation in retina images, and skin cancer segmentation. The new model provided a testing accuracy of 0.9424 with a higher AUC, which was 0.9419 on skin cancer segmentation.

Zhou et al. [40] developed a powerful architecture for medical image segmentation based on a deeply-supervised encoder-decoder network, which utilized skip pathways to reduce the differences between the two symmetrical ways. They achieved better performance on nodule segmentation in the low-dose CT scans of the chest. Isensee et al. [41] modified the structure of the U-Net slightly and improved the results by a Dice loss function. They trained the network with the regions of original images and additional training data. This method resulted in the highest Hausdorff Distances (95th percentile) of 6.03 for the whole tumor on the test data.

3. System model and definitions

We propose an RBCDS based on CNNs and multimode MR images to simulate the real clinical diagnostic process. The complete architecture of RBCDS is shown in Figure 1. The proposed classification model is modified with SPP and PCA. We compare the effects of different dimensionality reduction ratios on the results. Some data augmentation techniques, transfer learning, and fine-tuning methods are also used in this study. This model achieves the highest accuracy rate of 94.6%. For the optimized segmentation method, we compare the performance by evaluating the DC values of experiments. Apart from this, the classification network is used to initialize the segmentation network. Here we achieve an average DC value of 0.867.

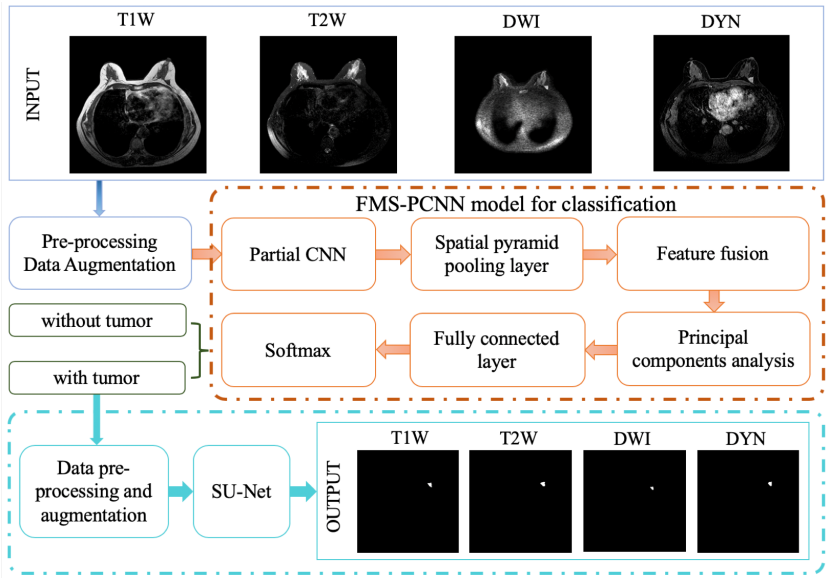

Figure 1. The proposed RBCDS. *Partial CNN*: the layers of network before the last pooling layer.

As a comparison, the traditional classification method commonly used is shown in Figure 2. This approach is a common process in the current diagnostic process.

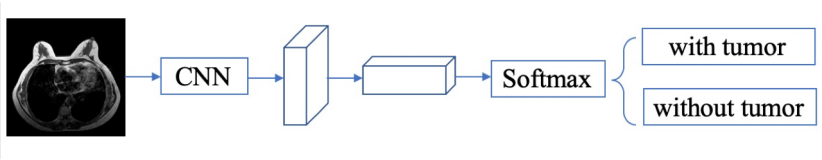

Figure 2. A classification process using a single-sequence image.

In our study, our MRI dataset includes T1W, T2W, DWI and DYN of 90 breast examinations from high-risk patients. The dataset is built at the radiology department of the Tianjin Center Obstetrics and Gynecology Hospital. In this work, 7000 images (2500 images with tumors and other images without tumors) are used for classification, and 2450 images with tumors are used for segmentation. The images are divided into training, verification, and test by 8:1:1. Radiologists use a 1.5 T scanner (PHILIPS MR Systems Achieva, version 3.2) for bilateral MRI.

The specific imaging methods of the four types of images are exhibited as following: T1W_TSE SENSE data are obtained in the transverse plane, scanning sequence is the spin-echo (SE) for 4.8 mm slice thickness, having 512×512 resolution, and 100×100 mm field of view (FOV), with 10 ms echo time (TE), and 456.6 ms repetition time (TR). T2W_SPAIR data are obtained in the transverse plane, using the SE for 4.8 mm slice thickness, each slice image has a 352×352 resolution, and 100×100 mm FOV, with 90 ms TE, and TR is 4858.8 ms. DWI_SSh SENSE data are obtained in the transverse plane, scanning sequence is SE for 4.8 mm slice thickness, 320×320 resolution, and 100×100 mm FOV, TE is 68.4 ms, and TR is 5869.2 ms. DYN_eTHRIVE+C data are obtained in the transverse plane and sagittal plane, using the gradient recalled (GR) for 340 mm slice thickness, 352×352 resolution, and 100×100 mm FOV, with 3.4 ms TE, and 6.9 ms TR.

For the segmentation task, images are preprocessed by a semi-automated data labeling system. A code written in Python is used for punctuation and connection to denote the boundaries of tumors in the images. The manual segmentation has been performed by three experienced radiologists, which can avoid the inaccuracy of decisions by a single radiologist. The radiologists used the code that we wrote to open each image containing the tumor and record the points on the edge around the tumor with the mouse. After the marking is completed, the completion button is triggered, and the tumor is displayed in a single image.

4. Our proposed classification model in breast cancer detection

In this study, images from four imaging modes are used to form the data set. In order to enable the network to process all images smoothly, we use SPP to optimize the network structure. The FMS-PCNN is also optimized by PCA. We make four comparative studies with different networks to verify that the proposed model is effective. The analyses intend to prove whether our proposed model is superior to the decision-level fusion method in every network. We also demonstrate the effectiveness of transfer learning [42] in our study through experiments. The algorithm of the classification task in breast cancer detection is explained in Algorithm 1. The features that entered into the SPP layer

are extracted by partial CNN in Figure 1. SPP plays a role in organizing features and learning them. Through convolution and pooling, our model can grasp important and detailed features of images. The extracted features are the basis of the following steps.

Algorithm 1 Process of the FMS-PCNN for classification task.

Require: preprocessed and augmented four-mode MR images with original sizes
1: **feature extraction:** extract features from processed images by partial CNNs in Figure 1
2: **spatial pyramid pooling:** replace the last pooling layer in the CNNs to avoid resizing images
3: **feature fusion:** fuse features output by last layer
4: **principal components analysis:** remove redundant information from the fused features
5: **fully connected layer**
6: **softmax**
Ensure: The classification results of the images

4.1. Preprocessing

Image quality directly affects the design and accuracy of the classification algorithm. Therefore, in addition to method optimization, preprocessing technology also plays an important role in the whole project.

We apply two standard preprocessing steps to the images. The entire image contains invalid information that interferes with the observed objects. We segment effective area related to the tumor from each image and then normalize the image by global contrast normalization (GCN), which is also named zero-centering. In general, normalization can eliminate or weaken the influence of some factors or transformation on some properties of the image. We remove the lung areas, leaving the breast tissue, the subphrenic region and the wall of the thoracic cavity that often adhered to the tumor. GCN attempts to deal with changes and differences brought by external conditions like illumination levels. We compute the mean of intensities for each image and subtract it from each pixel of the image. The new tensor of the normalized image is defined as

$$M'_{i,j} = M_{i,j} - \bar{M}, \tag{4.1}$$

where $M_{i,j}$ denotes the tensor of an image and $\bar{M}$ stands for the mean of the intensities of $M_{i,j}$. The image M satisfies $M \in \mathbb{R}^{r\times r}$. $\bar{M}$ is defined as:

$$\bar{M} = \frac{1}{r^2}\sum_{i,j} M_{i,j}. \tag{4.2}$$

4.2. Data augmentation

Data augmentation in computer vision is the artificial introduction of prior knowledge for visual invariance. Data augmentation is one of the commonly used skills in deep learning. It is mainly used to increase the amount of training data and make the dataset as diverse as possible so that the training model will have stronger generalization ability.

We augment the dataset by rotating the images 90, 180 and 270 degrees, and then applying random horizontal flip. Multiple patches are selected in the images. For baseline networks without SPP and

PCA, the images are resized to the appropriate input sizes for different networks. Four types of images are grouped together according to their imaging sequence and position. Finally, images are labeled with 0, 1, as the dichotomous variable. The label 0 means that there is no tumor in the image, the label 1 stands for the opposite.

4.3. Experimental details

In the process of image classification, we use five CNNs. Different experimental methods are employed in each network to examine the superiority of FMS-PCNN. In this way, we can avoid experimental specificity due to the particularity of specific structures.

4.3.1. Decision-level fusion

As shown in Figure 2, features of each single image are extracted through CNN. Next, they are forwarded to the softmax layer for two-category classification, and then the classification results of four images are summarized. If more than half of the images suggest the tumor in one group, the corresponding subject is considered to have a tumor, otherwise, there is no tumor. This method does not introduce the deep correlation of different imaging modes. Meanwhile, because of the complexity and diversity of the mammary glands, different effects appear across various types of images, leading to confusion in the observation of tumors.

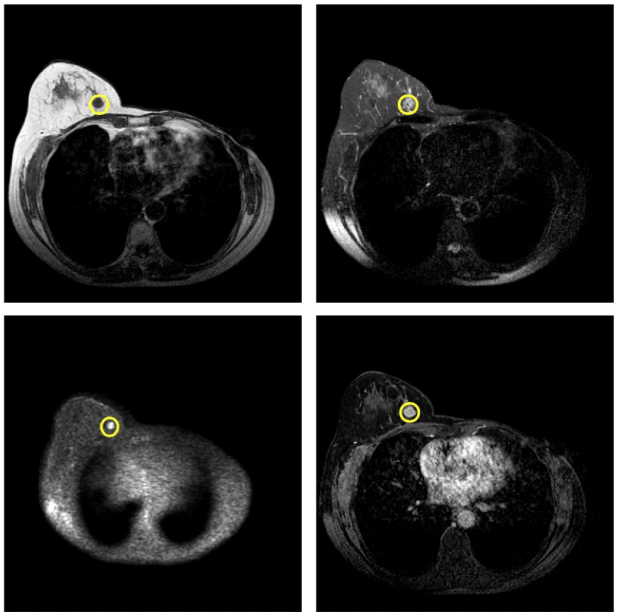

Figure 3. Different appearances of the same tumor in the same scan location from T1W, T2W, DWI and DYN of the same subject's breast.

The differences between four types of images in the same position can be seen in Figure 3. The images from left to right in the first row are T1W and T2W, in the second row are DWI and DYN. The tumors marked with yellow circles in DWI and DYN are distinguished from surrounding tissue, but they are not clear in T1W and T2W.Both images are from the same location on the same subject's

breast. In DWI, the breast tumor typically appears more intense to surrounding tissues [43]. This observation suggests that the radiologists may not be able to identify the tumor with a single-sequence image. Thus, we propose the detection model with connection of four sequential images. Among the numerous MR imaging sequences, we choose these four sequences for the following reasons: 1) a single image can not determine the diagnostic results; 2) radiologists often use four imaging modes in Figure 3 to analyze whether a subject has breast cancer. Combining multiple images for diagnosis can simulate the real diagnostic process more reliably.

4.3.2. SPP-networks

The popular networks are ended with pooling layers and fully connected layers followed by a C-way softmax as output, where C is the number of classes. They require a fixed size of an input image. Cropping and wrapping become the main methods to produce images of arbitrary sizes. Changing sizes of the input images may cause image losses and affect the performance of classification models. SPP layer [21] in the CNNs allows networks to take images of any size as input. It can generate fixed-length representations of features and maintain spatial information by pooling in local spatial bins. The number of bins is fixed regardless of different image sizes. Max pooling is used in SPP.

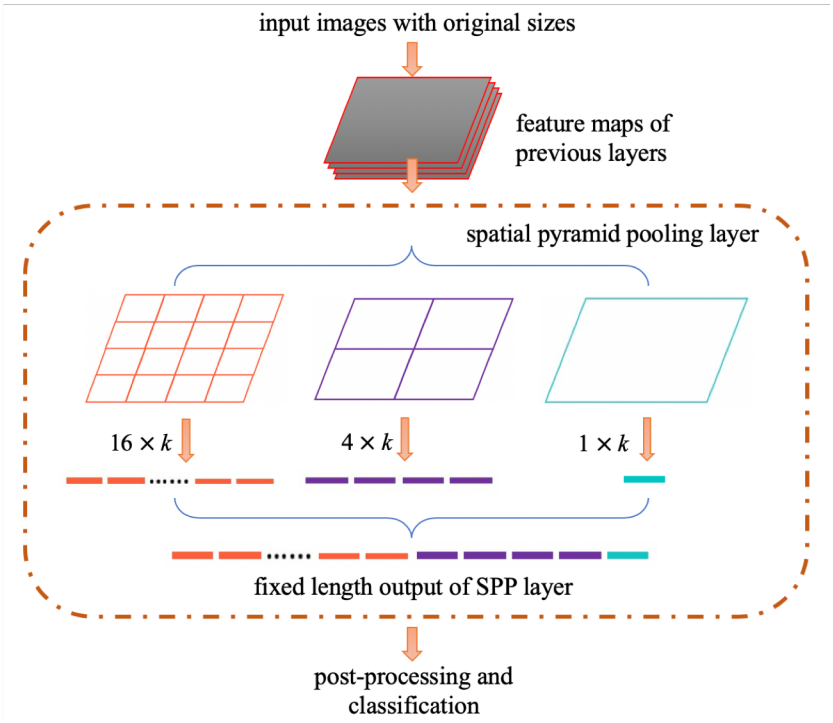

Figure 4. A network with a 3-level spatial pyramid pooling layer.

We replace the last pooling layer with an SPP layer. An SPP layer has an x-level pyramid pooling layer. Figure 4 shows a network with a 3-level SPP layer, which is applied in our model. Let k denote the number of filters in the last layer before the SPP layer and N denote the total number of bins. The outputs of the SPP layer are $k \times N$-dimensional features. If the feature map of the last layer before the

SPP layer is $m \times m$ and a pyramid level of all SPP layers is $n \times n$. The size of each bin is computed by

$$window = \lceil m/n \rceil, \tag{4.3}$$

and the corresponding stride is computed by

$$stride = \lfloor m/n \rfloor. \tag{4.4}$$

For example, when we use a 3-level SPP layer and the feature map of the last layer is 7×7, $pool_{4\times4}$ will have bins with size of 2×2 and stride of 1.

4.3.3. Feature fusion

Images produced by a single imaging method cannot be used as the only basis for diagnosing a patient. For the classification task, we combine the features of the four types of images to achieve better classification performance. As shown in Figure 1, in the classification process, after images of four sequences pass through the neural networks, the learned features are fused to form more representative features. This step can deeply take advantage of all the features from the parameters of four sequences and overcome the contingency of a single imaging mode. The n-th feature of the images output by previous layers is defined as t_n. We fuse the other features with t_n and compute the new n-th feature of images by

$$t_n^{'} = \sum_{c=1}^{4} \varphi^{xy} \left(K_S t_c\right), \tag{4.5}$$

where t_c is the feature of the c-th image in an imaging mode, K_S can transform the feature to a lower dimension, the maximum value of x is 4. The weight of linear sum is φ^{xy}, denoting the impact from other images, which can be computed as

$$\varphi^{xy} = \frac{exp\left(\varphi_Q^{xy}\right)}{\sum_k exp\left(\varphi_Q^{ky}\right)}. \tag{4.6}$$

The weight φ_Q^{xy} is computed as dot product and can be calculated as

$$\varphi_Q^{xy} = \frac{dot\left(K_U t_c, K_V t_c\right)}{\sqrt{d_k}}, \tag{4.7}$$

where K_U, K_V are matrices with similar function like K_S. And d_k is the dimension of the transformed feature. For the n-th feature, there are four transformed features, $t_n^{'(1)}$, $t_n^{'(2)}$, $t_n^{'(3)}$ and $t_n^{'(4)}$. Finally, all the $t_n^{'}$ are concatenated together and the n-th feature $t_n^{''}$ is rephrased as

$$t_n^{''} = t_n + Concat\left[t_n^{'(1)}, t_n^{'(2)}, t_n^{'(3)}, t_n^{'(4)}\right]. \tag{4.8}$$

After transformation, four new features $t_1^{''}$, $t_2^{''}$, $t_3^{''}$ and $t_4^{''}$ can be concatenated as a new feature [44]. The features corresponding to the four imaging parameters are learned by networks. We first combine features that belong to different imaging modes to produce an output. The different outputs, $t_n^{'(1)}$, $t_n^{'(2)}$, $t_n^{'(3)}$ and $t_n^{'(4)}$ are then combined with the original feature t_n. Lastly, all the calculated features of four modalities are concatenated as the final feature for image classification. On the basis of combining multiple sequence parameters, the classification model can be more reliable.

4.3.4. Dimensionality reduction by PCA

Differences between parameters of multiple sequences can be combined by feature fusion. But the similarities among multi-parameter modes may lead to repetitive information. In other words, fused features may contain redundant information, which could weaken the model's performance in the experiments. PCA [22, 45] can remove the information redundancy. Suppose that the feature needs to be reduced from a-dimension to b-dimension: PCA first finds b a-dimensional vectors, and then projects the features into b-dimensional space. It finally ensures that the projection error is small enough. Let $F = \left(f^{(1)}, f^{(2)}, \ldots, f^{(z)}\right)$ denote all the original components. We can compute new feature $g^{(i)}$ by

$$g^{(i)} = U_{\text{reduce}}^{T} \cdot f^{(i)}, \tag{4.9}$$

where $f^{(i)}$ denotes the i-th vector and U_{reduce} stands for a matrix of retained singular vectors. They can be calculated by

$$f^{(i)} = \frac{f^{(i)} - \mu^{(i)}}{s^{(i)}}, \tag{4.10}$$

$$U_{reduce} = \left(u^{(1)}, u^{(2)}, \cdots, u^{(b)}\right). \tag{4.11}$$

In the above formulas, $\mu^{(i)}$ is the mean of feature i and $s^{(i)}$ is the standard deviation of feature i. And $u^{(1)}, u^{(2)}, \cdots, u^{(b)}$ are the retained singular vectors. After the feature is normalized, the covariance matrix of it can be expressed as

$$\Sigma = \frac{1}{z} \sum_{i=1}^{z} \left(f^{(i)}\right)\left(f^{(i)}\right)^{T} = \frac{1}{z} \cdot F^{T} F. \tag{4.12}$$

Finally, we define the feature statistics rate (FSR) as the ratio of the number of principal components retained by PCA to all the original components. According to the proportion, we can obtain new features with less redundant information for classification. We first select FSR with values of 50, 70 and 90% in the experiments. Setting FSR to 90% means that the top 90% of principal components are selected. We secondly set FSR to 80% according to the highest accuracy of the above experiments. Preserving the most representative features can reduce the computational complexity and improve classification accuracy.

In the training, we select ResNet-50 [29], AlexNet [12], VGG 16 [31], Inception V3 [46] and DenseNet [47] as basic networks. All the networks are trained using the stochastic gradient descent (SGD) method. We use a batch size of 32 for 100 epochs. The weight decay is set to 1e-5 and a Nesterov momentum [25] is set to 0.9. The initial learning rate is set to 0.01 and is divided by 10 times at 50 and 75% of the total number of training epochs. We evaluate the models by classification accuracy and the AUC of the best model.

5. The segmentation model in RBCDS

For the segmentation task in our RBCDS, we introduce the SU-Net. We use the sub-pixel convolution in the upsampling process of U-Net [24] and the classification network to initialize the

SU-Net. Based on FCN, U-Net is modified and extended to process fewer training images and produce more accurate segmentation. It is a completely symmetrical architecture that only uses the valid part of each convolution. The skip connection promotes the gradient flow of the network. The algorithm of segmentation method [23] used in RBCDS is shown in Algorithm 2.

Algorithm 2 Process of the segmentation model in our RBCDS.

Require: preprocessed and augmented four-mode MR images with tumors

1: **encoding path:** extract features from images by downsampling process in contracting path

2: **decoding path:** use sub-pixel convolution method to optimize the upsampling process in expansive path

Ensure: The segmented images

5.1. Data preprocessing and augmentation

The images, including major areas related to the tumors, are normalized by GCN. Normalization can avoid image differences caused by light conditions and subtle human operation differences. We augment the data by rotating the images 90, 180 and 270 degrees. The random horizontal flip is also adopted for the segmentation task. We resize all the images to 256×256 as input to the segmentation network.

5.2. Experimental methods

The U-Net [24] model is improved for the segmentation task. Since the upsampling process of U-Net only uses a simple method called the nearest neighbor method to copy the neighborhood points. This method for upsampling is too rough, especially for the segmentation of pixel-level tasks. A rough upsampling method may lead to less clear edges of the segmentation results.

For the above reasons, we specifically use the sub-pixel convolution method to optimize the upsampling process in U-Net. The sub-pixel convolution is a standard convolution that can learn an array of upscaling filters to upscale the low resolution (LR) feature maps into the high resolution (HR) output [25]. In image processing for medical imaging, the recovery of an HR image from its LR counterpart refers to as super-resolution (SR) [25, 48]. The sub-pixel convolution method can contribute to better localization of edges through image upsampling and it can remove the blocking artifact in image SR task. The trained model in the classification is used to initialize the left half of SU-Net (namely the encoder process).

All the networks are trained using the SGD method. The batch size of 16 for 80 epochs is employed. The network uses a non-linear transformation (ReLU). The initial learning rate is set to 0.001, divided by 10 times at 40 and 60 epochs. The weight decay is set to 1e-5 and Nesterov momentum [25] is set to 0.9, which is the same as the classification model.

We evaluate the performance of the SU-Net by dice coefficient, which can reflect the similarity between ground truth and segmentation result of the model. The value of DC can be computed by

$$\mathrm{DC}\left(X_{\mathrm{gt}}, X_{\mathrm{ans}}\right) = \frac{2\left|X_{\mathrm{gt}} \cap X_{\mathrm{ans}}\right|}{\left|X_{\mathrm{gt}}\right| + \left|X_{\mathrm{ans}}\right|}, \tag{5.1}$$

where X_{gt} and X_{ans} represent the ground truth and auto-segmentation of a given image.

6. Performance analysis

6.1. Classification results

This section describes the results of the FMS-PCNN model with five CNNs. The start-up training with the pre-trained model on ImageNet is also compared. The benefits of using multimode MR images for training are analyzed based on the experimental results. Using transfer learning and fine-tuning strategies can improve the classification accuracy.

Figure 5 shows the results using the decision-level fusion method. Figure 6 reveals better performance based on SPP, PCA, and fused features. Using three training strategies (transfer learning, fine-tuning and data augmentation) simultaneously improves the results. The data can strongly demonstrate that using SPP and PCA in the networks produces higher accuracy. This is due to the use of original-sized images and the reduction of superfluous information in the fused features. The accuracy achieves 92.6% without PCA. In the case of multiple images, each method uses four paths. The model does not use multiple fully connected layers containing a large number of parameters, but directly maps to the fully connected layers activated by classification at the end, which compensates for the parameter increase caused by fusion. Our best result of this step is: FMS-PCNN reaches the accuracy of 92.9% with FSR set to 50%.

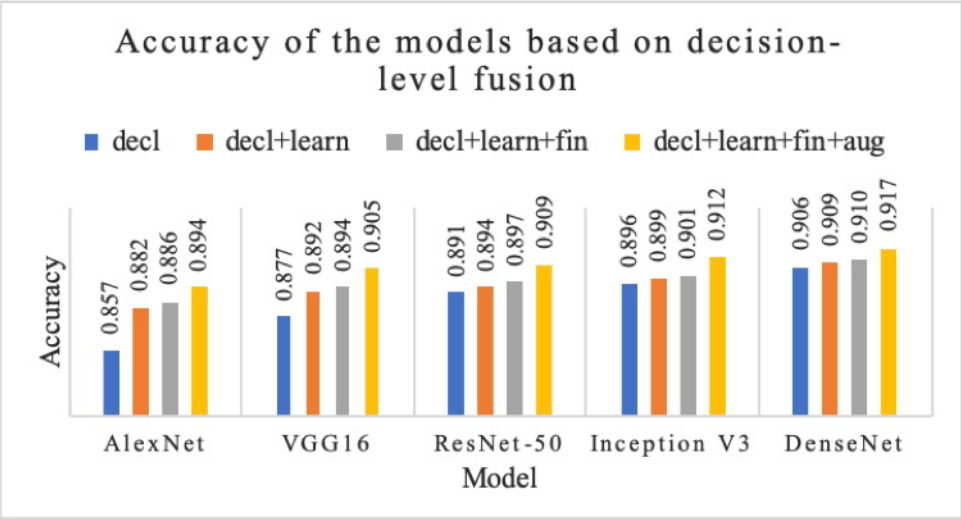

Figure 5. The comparisons of the results based on decision-level fusion using basic networks without SPP and PCA. *decl*: training networks using decision-level fusion method. *decl+learn*: training networks with transfer learning using decision-level fusion method. *decl+learn+fin*: training networks with transfer learning and fine-tuning using decision-level fusion method. *decl+learn+fin+aug*: training networks with transfer learning, fine-tuning and data augmentation strategies using decision-level fusion method.

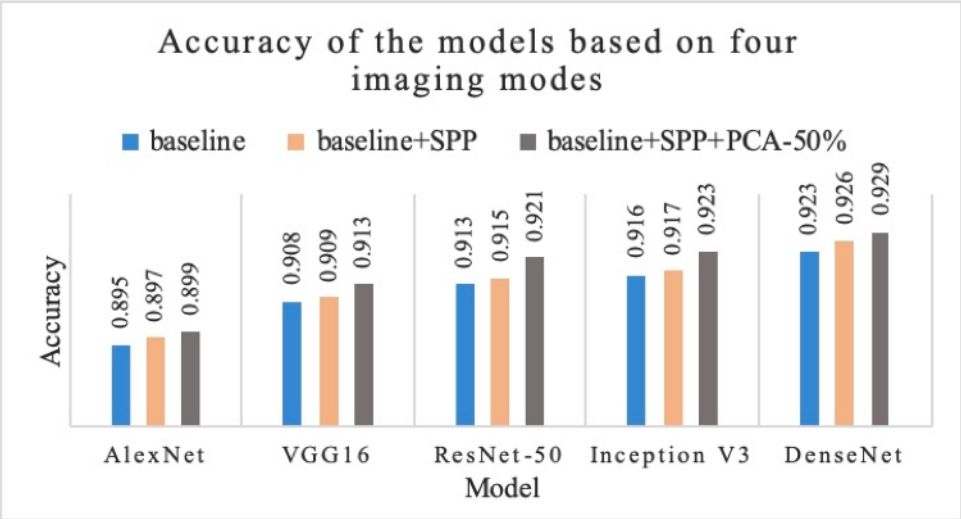

Figure 6. The comparisons of the results based on four imaging modes and feature fusion. *baseline*: training basic networks with three strategies (transfer learning, fine-tuning and data augmentation) using four-mode fusion method. *baseline+SPP*: adding SPP to baseline. *baseline+SPP+PCA-50%*: adding SPP to baseline and setting FSR to 50%.

Table 1 reveals the comparisons of classification accuracy values across different FSR values. Encouragingly, the accuracy improves to 94.6% when FSR is set to 80%, indicating redundancy between the raw fused features. However, when we further increase the FSR to 90%, the accuracy drops to 92.8%. And higher FSR results in greater computational complexity. The retained top 90% of components of all components still have invalid information for classification. If fewer features are retained, the classification accuracy will be lower because of the incompleteness of features. In brief, we should find the most appropriate proportion to achieve the best classification performance, instead of reducing the capacity of feature too much.

Table 1. The comparisons of classification accuracy across different values of FSR with DenseNet based on four sequence images. Each method is implemented with transfer learning, fine-tuning and data augmentation simultaneously.

FSR	50%	70%	80%	90%
Accuracy	0.929	0.939	0.946	0.928

Figure 7 shows that AUC of the best classification model achieves 0.928 with an accuracy of 0.946. In addition, we compare our results with other methods, as shown in Table 2. Multi-parameter images are the most important foundation for outstanding results. The other methods use dynamic contrast-enhanced MRI (DCE-MRI) and mammogram images. MR images are clearer than other images and can provide more details of the soft tissue. The feature fusion of multi-sequence images can greatly improve the classification accuracy and the AUC is also improved. At the same time, the use of multi-

sequence images is more consistent with the real diagnostic process.

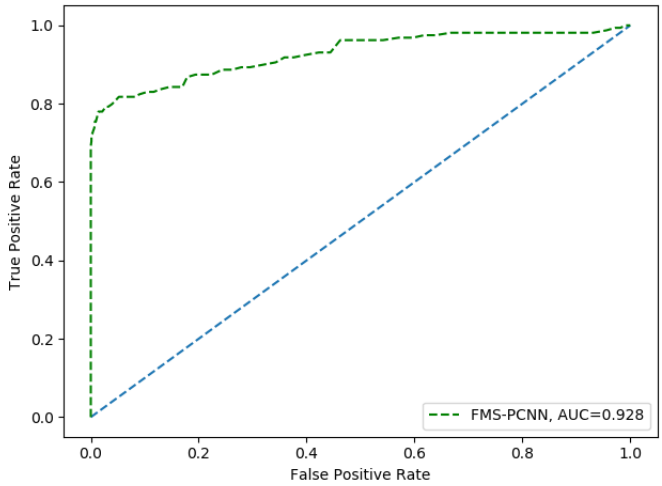

Figure 7. Classification performance. The green line is the ROC curve of the optimal classification model on the dataset. The corresponding AUC value is 0.928.

Table 2. The comparisons of classification results across different methods.

Method	Data	AUC	Accuracy
Haarburger C. [49]	DCE-MRI	0.890	0.810
Fung Fung Ting [50]	mammogram	0.901	0.905
Maicas G. [51]	DCE-MRI	0.900	–
Ours	T1W, T2W, DWI, DYN	0.928	0.946

6.2. Segmentation results

Figure 8 exhibits the DC values of using the basic U-Net, only using initialization, and using both initialization and SU-Net. The best mean DC value of segmentation achieves 0.867 by utilizing initialization and the proposed SU-Net model concurrently. Segmentation results can help clinicians further observe the lesions and formulate follow-up treatment plans.

To verify the effectiveness of the proposed approach, Table 3 compares the performance obtained by some literature proposals. The 2D Fuzzy C-Means (2D-FCM) [52], Fully Convolutional Networks (FCN) [53], and SegNet [54] are mainly compared. Our segmentation model obviously achieves better results.

Finally, the accurate segmentation process can be integrated with the classification model to form a system for the diagnosis of breast cancer. The multimode fusion method in the classification task prevents the confusion between observation and breast tumors and simulates the clinical diagnosis

process. The improved segmentation network provides more detailed tumor edge information. These results can help radiologists and clinicians make inspection reports and determine the therapeutic plans.

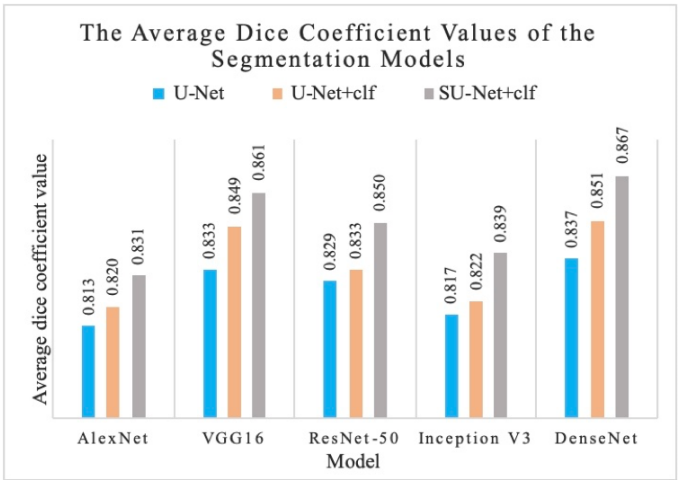

Figure 8. The comparisons of the results across basic U-Net and SU-Net. *U-Net+clf*: initializing U-Net with classification network. *SU-Net+clf*: initializing the SU-Net by classification network.

Table 3. Comparison of the proposed SU-Net approach with some proposals in segmentation of medical images.

Method	2D-FCM	FCN	SegNet	Ours
DC value	0.835	0.849	0.851	0.867

7. Conclusions and future works

We have proposed an RBCDS based on multimode MR images in this paper. Firstly we develop the FMS-PCNN model for image classification in the breast cancer diagnostic system. This model fuses MRI-based features to combine multiple parameters of four imaging modes. The model also avoids resizing images and retains the most representative features. Secondly, we optimize the upsampling process of U-Net and initialize SU-Net by classification network. The performance of the proposed model surpasses the traditional U-Net model. In the future, the segmentation results can be correlated with the blood flow information and density characteristics of the tumor area for diagnosis. Finally, the proposed classification and segmentation models form a complete robust diagnostic system, which can solve the confusing problem that tumors are difficult to distinguish. Instead of a single image, the system can diagnose four images, making the results more robust. This system can be used as an auxiliary system in the actual diagnostic process to provide information for clinicians.

In the future, more targeted and complex feature fusion methods and sequences can be used for

breast cancer detection and achieve excellent performance. With the development and innovation of imaging technology, the researchable images and imaging modes available will be more comprehensive. Additional parameters of other imaging sequences of MRI will be helpful in improving breast cancer detection methods and systems. Further research on the extent of disease and the presence of multifocal cancer in patients with invasive carcinoma and ductal carcinoma in situ (DCIS) will rely heavily on MR images.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No.61876131, U1936102) and the National Key R&D Program of China (No.2018YFC0806802).

Conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

1. R. L. Siegel, K. D. Miller, A. Jemal, Cancer Statistics, 2019, *Cancer J. Clin.*, **69** (2019), 7–34.
2. K. DeikeHofmann, F. Koenig, D. Paech, C. Dreher, S. Delorme, H. Schlemmer, et al., Abbreviated MRI Protocols in Breast Cancer Diagnostics, *J Magn. Reson. Imaging*, **49** (2019), 647–658.
3. H. Chougrad, H. Zouaki, O. Alheyane, Deep Convolutional Neural Networks for breast cancer screening, *Comput. Mehtod Prog. Biomed.*, **157** (2018), 19–30.
4. A Gubernmrida, M Kallenberg, R. M. Mann, R Mart, N Karssemeijer, Breast segmentation and density estimation in breast MRI: a fully automatic framework, *IEEE J. Biomed. Health Infor.*, **19** (2015), 349–357.
5. R. M. Mann, N. Cho, L. Moy, Breast MRI: State of the Art, *Radiology*, **292** (2019), 520–536.
6. C. D. Lehman, R. D. Wellman, D. S. M. Buist, Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection, *JAMA Int. Med.*, **175** (2015), 1828–1837.
7. T. Kyono, F. J. Gilbert, M. van der Schaar, MAMMO: A Deep Learning Solution for Facilitating Radiologist-Machine Collaboration in Breast Cancer Diagnosis, preprint, arXiv:1811.02661.
8. F. A. Maken, A. P. Bradley, *Multiple Instance Learning for Breast MRI Based on Generic Spatio-temporal Features*, 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.
9. Y. Lecun, Y. Bengio, G. Hinton, Deep learning, *Nature*, **521** (2015), 436–444.
10. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., *Going Deeper with Convolutions*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
11. Z. Fang, F. Fei, Y. Fang, C. Lee, N. Xiong, L. Shu, et al., Abnormal event detection in crowded scenes based on deep learning, *Multimed Tools Appl.*, **75** (2016), 14617–14639.

12. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.*, **25** (2012), 1097–1105.

13. J. Long, E. Shelhamer, T. Darrell, *Fully Convolutional Networks for Semantic Segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

14. Q. Hu, C. Wu, Y. Wu, N. Xiong, UAV Image High Fidelity Compression Algorithm Based on Generative Adversarial Networks Under Complex Disaster Conditions, *IEEE Access*, **7** (2019), 91980–91991.

15. Y. Xu, H. Yang, J. Li, J. Liu, N. Xiong, An Effective Dictionary Learning Algorithm Based on fMRI Data for Mobile Medical Disease Analysis, *IEEE Access*, **7** (2018), 3958–3966.

16. C. Wu, C. Luo, N. Xiong, W. Zhang, T. H. Kim, A Greedy Deep Learning Method for Medical Disease Analysis, *IEEE Access*, **6** (2018), 20021–20030.

17. D. Truhn, S. Schrading, C. Haarburger, H. Schneider, D. Merhof, C. Kuhl, Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-enhancing Lesions at Multiparametric Breast MRI, *Radiology*, **290** (2019), 290–297.

18. N. Chen, T. Qiu, X. Zhou, K. Li, M. Atiquzzaman, An Intelligent Robust Networking Mechanism for the Internet of Things, *IEEE Commun. Mag.*, **57** (2019), 91–95.

19. T. Qiu, J. Liu, W. Si, D. O. Wu, Robustness optimization scheme with multi-population co-evolution for scale-free wireless sensor networks, *IEEE Trans. Networking*, **27** (2019), 1028–1042.

20. T. Qiu, B. Li, W. Qu, E. Ahmed, X. Wang, TOSG: A topology optimization scheme with global-small-world for industrial heterogeneous Internet of Things, *IEEE Trans. Ind. Inf.*, **15** (2019), 3174–3184.

21. K. He, X. Zhang, S. Ren, J. Sun, Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, **37** (2015), 1904–1916.

22. N. N Mohammed, M. I. Khaleel, M. Latif, Z. Khalid, *Face Recognition Based on PCA with Weighted and Normalized Mahalanobis distance*, 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2018.

23. W. Lu, Z. Wang, Y. He, H. Yu, N. Xiong, J. Wei, *Breast Cancer Detection Based On Merging Four Modes MRI Using Convolutional Neural Networks*, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.

24. O. Ronneberger, P. Fischer, T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.

25. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, et al., *Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

26. F. A. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte, *Breast Cancer Histopathological Image Classification using Convolutional Neural Networks*, 2016 International Joint Conference on Neural Networks (IJCNN), 2016.

27. Y. Wen, K. Zhang, Z. Li, Y. Qiao, *A Discriminative Feature Learning Approach for Deep Face Recognition*, European Conference on Computer Vision, 2016.
28. Y. Gao, X. Xiang, N. Xiong, B. Huang, H. Jong Lee, R. Alrifai, et al., Human Action Monitoring for Healthcare Based on Deep Learning, *IEEE Access*, **6** (2018), 52277–52285.
29. K. He, X. Zhang, S. Ren, J. Sun, *Deep Residual Learning for Image Recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
30. Z. Wang, W. Lu, Y. He, N. Xiong, J. Wei, *RE-CNN: A Robust Convolutional Neural Networks for Image Recognition*, International Conference on Neural Information Processing, 2018.
31. K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, preprint, arXiv:1409.1556.
32. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Object Detectors Emerge in Deep Scene CNNs, preprint, arXiv:1412.6856.
33. B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, A. Oliva, *Learning Deep Features for Scene Recognition Using Places Database*, Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014.
34. S. Liu, J. Zeng, H. Gong, H. Yang, J. Zhai, Y. Cao, et al., Quantitative analysis of breast cancer diagnosis using a probabilistic modelling approach, *Comput. Biol. Med.*, **92** (2018), 168–175.
35. A. M. Sayed, E. Zaghloul, T. M. Nassef, Automatic classification of breast tumors using features extracted from magnetic resonance images, *Procedia Comput. Sci.*, **95** (2016), 392–398.
36. K. Drukker, R. Anderson, A. Edwards, J. Papaioannou, F. Pineda, H. Abe, et al., *Radiomics for Ultrafast Dynamic Contrast-Enhanced Breast MRI in the Diagnosis of Breast Cancer: a Pilot Study*, Medical Imaging 2018: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2018,
37. G. Maicas, G. Carneiro, A. P. Bradley, *Globally Optimal Breast Mass Segmentation from DCE-MRI Using Deep Semantic Segmentation as Shape Prior*, 2017 IEEE 14th International Symposium on Biomedical Imaging, 2017.
38. J. Zhang, A. Saha, Z. Zhu, M. A. Mazurowski, *Breast Tumor Segmentation in DCE-MRI Using Fully Convolutional Networks with an Application in Radiogenomics*, Medical Imaging 2018: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2018.
39. M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, V. K. Asari, Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation, preprint, arXiv:1802.06955.
40. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, *UNet++: A Nested U-Net Architecture for Medical Image Segmentation*, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018.
41. F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K. H. M. Hein. *No New-net*, Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 2018.
42. H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I Nogues, et al., Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, *IEEE Trans. Med. Imaging*, **35** (2016), 1285–1298.

43. N. Amornsiripanitch, S. Bickelhaupt, H. J. Shin, M. Dang, H. Rahbar, K. Pinker, et al., Diffusion-weighted MRI for Unenhanced Breast Cancer Screening, *Radiology*, **293** (2019), 504–520.
44. H. Hu, J. Gu, Z. Zhang, J. Dai, Y. Wei, *Relation Networks for Object Detection*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
45. X. Lu, X. Duan, X. Mao, Y. Li, X. Zhang, Feature Extraction and Fusion Using Deep Convolutional Neural Networks for Face Detection, *Math. Prob. Eng.*, **2017** (2017), 1376726.
46. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
47. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, *Densely Connected Convolutional Networks*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
48. W. Shi, J. Caballero, C. Ledig, X. Zhuang, W. Bai, K. Bhatia, et al., *Cardiac Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch*, Medical Image Computing and Computer-Assisted Intervention, MICCAI 2013.
49. C. Haarburger, M. Baumgartner, D. Truhn, M. Broeckmann, H. Schneider, S. Schrading, et al., *Multi Scale Curriculum CNN for Context-Aware Breast MRI Malignancy Classification*, Medical Image Computing and Computer Assisted Intervention, MICCAI 2019.
50. F. F. Ting, Y. J. Tan, K. S. Sim, Convolutional neural network improvement for breast cancer classification, *Expert Syst. Appl.*, **120** (2019), 103–115.
51. G. Maicas, A. P. Bradley, J. C. Nascimento, I. Reid, G. Carneiro, *Training Medical Image Analysis Systems like Radiologists*, Medical Image Computing and Computer Assisted Intervention, MICCAI 2018.
52. S. Marrone, G. Piantadosi, R. Fusco, A. Petrillo, M. Sansone, C. Sansone, *Breast Segmentation Using Fuzzy C-Means and Anatomical Priors in DCE-MRI*, 2016 23rd International Conference on Pattern Recognition (ICPR), 2016.
53. J. Long, E. Shelhamer, T. Darrell, *Fully Convolutional Networks for Semantic Segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
54. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, *IEEE Trans Pattern Anal. Mach. Intell.*, **39** (2017), 2481–2495.

AIMS Press