Knowledge-Based Systems 208 (2020) 106465

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Neural networks model based on an automated multi-scale method for mammogram classification

Lizhang Xie [a], Lei Zhang [a,*], Ting Hu [a], Haiying Huang [b], Zhang Yi [a]

[a] Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, 610065, PR China
[b] Information Management Department of West China Second University Hospital, Sichuan University, Chengdu, Sichuan, PR China

ARTICLE INFO

Article history:
Received 15 October 2019
Received in revised form 10 July 2020
Accepted 2 September 2020
Available online 19 September 2020

Keywords:
Breast cancer
Mammogram classification
Small lesions
Convolutional neural networks
Multi-scale feature

ABSTRACT

Breast cancer is the most commonly diagnosed cancer among women. Convolutional neural networks (CNN)-based mammogram classification plays a vital role in early breast cancer detection. However, it pays too much attention to the lesions of mammograms and ignores the global characteristics of the breast. In the process of diagnosis, doctors not only pay attention to the features of local lesions but also combine with the comparison to the global characteristics of breasts. Mammogram images have a visible characteristic, which is that the original image is large, while the lesions are relatively small. It means that the lesions are easy to overlook. This paper proposes an automated multi-scale end-to-end deep neural networks model for mammogram classification, that only requires mammogram images and class labels (without ROI annotations). The proposed model generated three scales of feature maps that make the classifier combine global information with the local lesions for classification. Moreover, the images processed by our method contain fewer non-breast pixels and retain the small lesions information as much as possible, which is helpful for the model to focus on the small lesions. The performance of our method is verified on the INbreast dataset. Compared to other state-of-the-art mammogram classification algorithms, our model performs the best. Moreover, the multi-scale method is applied to the networks with fewer parameters that can achieve comparable performance, while saving 60% of the computing resources. It shows that the multi-scale method can work for both performance and computational efficiency.

1. Introduction

Nowadays, breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer death among women globally [1,2]. In China, the average age at diagnosing for breast cancer is decreasing, and the number of breast cancer patients is increasing annually [3]. Early detection and early treatment are crucial to lowering the death rates of breast cancer [4]. Fortunately, some imaging techniques, such as mammogram screening and ultrasound examinations, are beneficial to early breast cancer detection. Mammogram screening has several advantages, including easy operating, affordable price, and significantly better imaging results of calcified breast cancer, which makes diagnosis easier. For these reasons, mammogram screening has become the most popular technology to diagnose breast diseases in women over 40 years old [5].

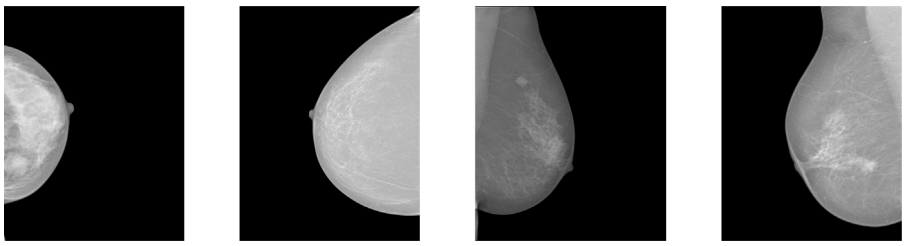
In general, the clinicians of the radiology department scan the left and right breasts of patients from the craniocaudal (CC) and mediolateral oblique (MLO) views. Hence, mammogram screening usually contains four images, from left-CC, right-CC, left-MLO, and right-MLO views [6], respectively. Fig. 1 shows the example of mammogram images with the four views.

Mammogram images reveal lots of information about the breast, including the density, shape and size of the breast, as well as the suspected lesions such as calcifications and masses, etc. The mentioned information is beneficial to well-trained doctors for making a preliminary diagnosis. Usually, some hospitals have mammogram scanning equipment and operating clinicians but are in a lack of well-trained doctors. Therefore, the development of trained doctors-level computer-aided diagnosis (CAD) systems to diagnose mammogram images will be of great benefit [4].

Mammogram classification is the most critical part of the CAD systems. Traditional mammogram classification methods typically focus on constructing hand-crafted features to describe the characteristics of breast cancer [7–9]. These features must be carefully designed based on the region-of-interest (ROI) annotations and the doctor's experience. The defects of these approaches can be summarized as follows. The first, ROI annotations require

* Corresponding author.
E-mail addresses: xielizhang1994@stu.scu.edu.cn (L. Xie), leizhang@scu.edu.cn (L. Zhang), huting@stu.scu.edu.cn (T. Hu), xuxu2110@163.com (H. Huang), zhangyi@scu.edu.cn (Z. Yi).

https://doi.org/10.1016/j.knosys.2020.106465
0950-7051/



(a) left-CC. (b) right-CC. (c) left-MLO. (d) right-MLO.

Fig. 1. Mammogram samples from fours views.

trained doctors to make pixel-level annotations on the image, which consumes a lot of effort. The second, the hand-crafted features are limited by medical knowledge about breast cancer and the precision of ROI annotations [10], which means that hand-crafted features may not be able to describe the characteristics of breast cancer precisely. The third, from the results reported [7–9], the classification and generalization performances of traditional methods are not good enough for clinical use.

Recently, the CNN-based method has achieved great success in computer vision and related research fields for its powerful feature extraction ability. The CNN method has high potential in mammogram image analysis, but the following challenges remain [11–17].

(1) CNN tends to achieve great performance with a large amount of data. But mammogram datasets have much fewer samples than natural image datasets. For instance, INbreast [18] is an open dataset of mammogram images and only has 410 images. In contrast, ImageNet [19] has 14,197,122 images and MSCOCO [20] has 82,783 images. Therefore, it is difficult to train a CNN model that performs well on the mammogram dataset.

(2) The dimensions of mammogram images usually are higher than those of natural images. The sizes of the images in INbreast are about 2000 $\times$ 3000 pixels, whereas those in ImageNet and MSCOCO are about 300 $\times$ 300 pixels. CNN usually resizes the input images to a fixed size, such as 299 $\times$ 299 or 224 $\times$ 224. For the raw mammogram images are too large, some information of lesions may be lost by directly resizing, which will result in that CNN cannot learn from these lesions.

(3) As can be seen in Fig. 1, there is a significant amount of redundant region in the raw mammogram images. The lesions only exist in the breast, and the breast region only occupies less than half of the whole picture. The redundant areas are not only useless for classification but also interfere with model training.

(4) Doctors combine the features of the whole breast and the suspected lesions in mammogram images for the diagnosis, while CNN usually relies on its powerful feature extraction ability to classify the natural pictures. Directly applying the CNN model to mammogram classification is difficult to achieve good performance [13,14,21].

(5) At present, current CNN models achieve excellent performance with ROI annotations-assisted training. However, ROI labeling is expensive and not easy to obtain [22,23]. Moreover, the performance of a CNN that training without ROI annotations is not good enough for practical application [12,21]. Improving the performance of classification models without ROI annotations is full of challenges.

To deal with the mentioned challenges, this paper proposes a multi-scale CNN model for mammogram classification. The model consists of the breast region segmentation (BRS) module, the feature extraction module, the multi-scale feature module, and the classifier module. The images preprocessed by the BRS module contain fewer non-breast (2) and (3) with a practical solution. To make challenge (1) less challenging, random transforms are performed on the input images firstly. Next, the dense connection mechanism of DenseNet can avoid overfitting to some degree, which is helpful for the training of the networks on the small dataset. So the pre-trained DenseNet is used for the feature extraction module. It learns from the mammogram images to generate feature maps representing the calcifications and masses. After that, the multi-scale module fuses these feature maps and generates new ones at three scales. The three scale feature maps represent the information of the whole breast and the suspected lesions, which offers challenge (4) a feasible way. Its performance is comparable to the model with ROI annotations-assisted training, and this can address challenge (5). Finally, this paper uses a fully connected layer as the classifier. The classifier collects multi-scale instead of single-scale feature maps, which helps the classifier work better.

The main contributions of this paper are summarized as follows.

(1) Based on the analysis of the doctor's diagnostic process and the characteristics of mammogram images, an automated multi-scale CNN model for mammogram classification is proposed in this paper. This method only requires original mammogram images and the corresponding category labels (without ROI annotations) and gets the state-of-the-art classification performance on the INbreast dataset. The proposed model saves a lot of efforts of labeling and makes the model more comfortable to apply.

(2) The multi-scale module is proposed to generate feature maps at three scales, which provides the model with the information of global breast and local lesions instead of only focuses on the local lesions. Moreover, considering the finite computational ability, the DenseNet [24] is replaced with the MobileNet [25], which can save 60% of computing resources while maintaining comparable performance.

(3) The BRS module is proposed to preprocess the raw mammogram images based on the characteristics of mammogram images. The processed images contain fewer non-breast pixels and the image size is small, which helps the model to focus on the breast region for better training.

2. Related work

In recent years, neural networks have been swiftly developed in theory and application [26–30], especially CNN. Since AlexNet [31] won the championship in the ILSVRC2012 challenge, CNN has attracted widespread attention from researchers. Subsequently, the structures of CNN, such as VGG Net [32], Inception Net [33], ResNet [34] and DenseNet [24], were designed to be deeper, which resulted in better feature extraction capabilities. The development of these structures has promoted the application of CNN in computer vision [35,36], especially for medical image analysis [37–43]. Wang et al. [37] constructed automated retinopathy of prematurity detection model using a CNN-based method. This approach built identification and grading models for two-stage learning and achieved great performance. Esteva et al. [39] trained a CNN model to classify images of skin lesions

as benign or malignant skin cancers. This model has achieved the accuracy of board-certified dermatologists. Zhu et al. [42] presented an automated lung CT cancer diagnostic system called DeepLung, which consists of nodule detection and classification components. It surpassed the performance of experienced doctors based on image modality. These CNN methods have achieved advanced performance in the field of medical image analysis.

At present, CNN-based methods have also achieved many breakthroughs in the mammogram classification. Li et al. [22] proposed an end-to-end networks with fully convolutional layers for mammogram classification. This structure is implemented with a fully convolutional layer and a removed pooling layer. RF+CNN [23] uses CNN to extract features from mammogram images and then uses a random forest as a classifier. These methods generally use mammogram images, ROI annotations, and category labels, and achieve good performance. However, ROI annotations require a trained doctor to create pixel-level annotations on the image, which consumes a lot of human resources.

Hence, some works have also explored CNN methods without ROI annotations for mammogram classification. For example, Agarwal et al. [12] used a pre-trained CNN model on ImageNet and fine-tuned the parameters and weights for mammogram classification. This is a relatively mainstream transfer learning approach and also very suitable for medical datasets. The sparse MIL method [21] was an end-to-end trained deep multi-instance networks for mass classification. It combined the sparse characteristics of breast lesions and multi-instance learning methods. Li et al. [44] used U-Net, which is pre-trained with medical images to improve the performance of breast mass segmentation. Other studies [13,14] combined MLO and CC views to diagnose breast cancer. However, the MLO and CC views of some datasets are not uniformly labeled but instead labeled separately. Domingues et al. [15] improved model performance through preprocessing the mammogram images, which is also an effective way to enhance the performance of models in general.

The above methods improve mammogram classification performance by various techniques, including networks architecture innovation, data preprocessing, and transfer learning. However, none of these approaches achieve comparable performance to the method with the ROI annotations-assisted training. It implies that ROI annotations have a substantial influence on the classification performance. In the study of classification with ROI annotations-assisted training, neural networks tend to directly learn the difference between the ROI and non-ROI annotations region. It is a typical image classification that CNN is particularly good at. In the absence of ROI annotations, CNN is required to master skills in the diagnosis of mammogram lesions.

Fig. 2 shows the diagnostic process of the doctor and the proposed model. A well-trained doctor's diagnosis usually involves four steps. Firstly, the doctor looks only at the breast region in the image and collects the global features such as the density, shape and size of the breast. Next, he or she looks for suspected lesions regions with high breast density. And then, focusing on these suspected lesions regions, the doctor identifies them as calcifications or masses, which is depending on their shape and size, etc. Finally, since the features of the breast vary from person to person, the doctor will compare the features of the whole breast and the suspected lesions. The process of comparison mainly focuses on the following two parts. One is to compare the density of the suspected regions and the surrounding regions. The other is to compare the size and shape of the suspected regions as well as the whole breast. According to the comparison results, the doctor will make a preliminary diagnosis combined with their own experience.

Based on the above description, two characteristics can be summarized as follows. The first, the doctors ignore the non-breast regions of the image, because these regions do not affect the diagnosis. The second, the doctors combine with the features of the whole breast and the suspected lesions to diagnose, and the works mentioned are usually only focused on the characteristics of local lesions.

The convolutional layer generates feature maps to represent the characteristics of the local region. Because the convolution kernel is tiny, so in general, it is a small region. The pooling layer merges these features and lowering the dimensions of the feature maps. Therefore, the pooling layer will inevitably lose some subtle information. The fully connected layer is used as a classifier. The above process is reasonable for the application of natural image classification. For example, as long as CNN captures a cat's features, it can support the idea that a cat is in the picture. By contrast, the process is inconsistent with the process of classifying mammogram images.

For these reasons, a multi-scale module is employed to improve CNN's classifying process of mammogram images. Like a doctor, it can focus on the global view and the local lesions' information for diagnosis. As shown in Fig. 2, The large-scale, mesoscale and small-scale features correspond to the processes of focusing on the whole breast, looking for suspected lesions in the breast and focusing on them, respectively. After collecting the features of the whole breast and the suspected lesions, a

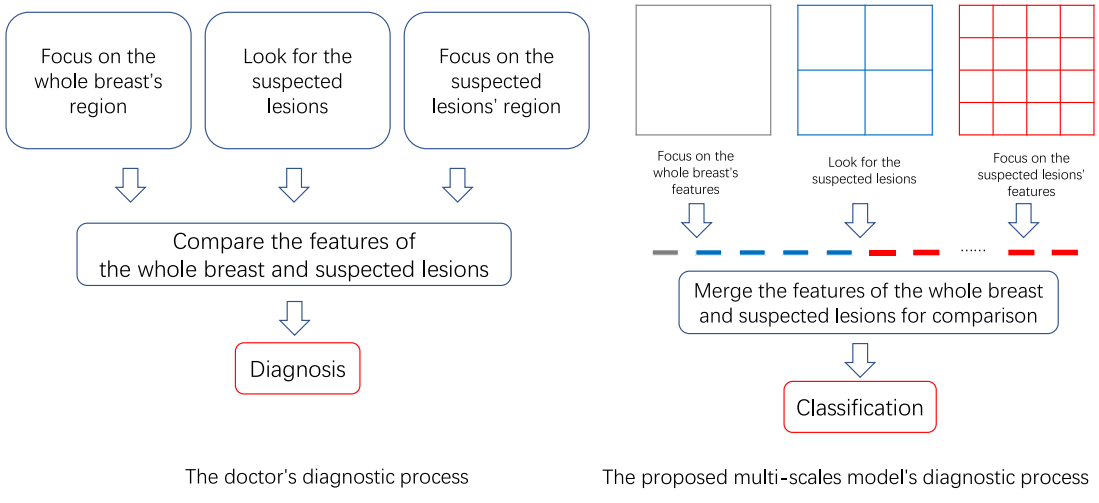


Fig. 2. The diagnostic process of doctor and the proposed model.

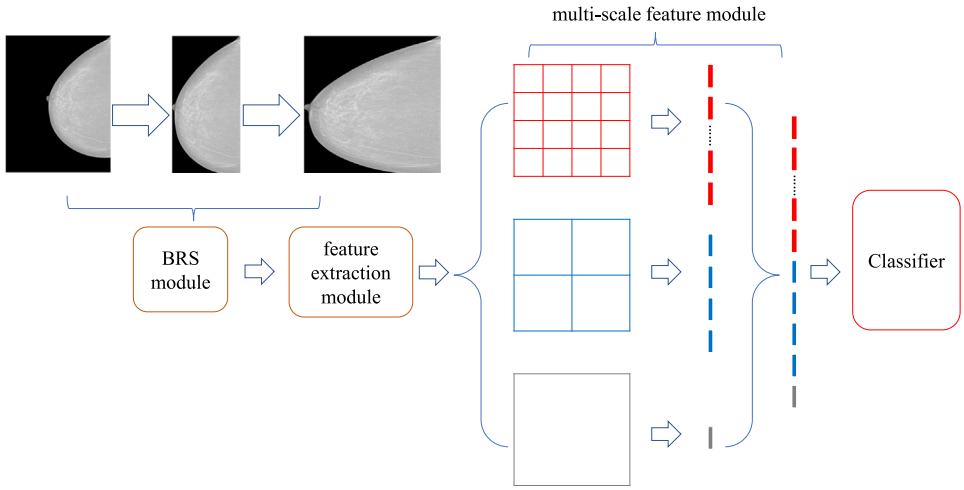


Fig. 3. The structure of the proposed model.

classifier was applied to diagnose. In this paper, the spatial pyramid pooling method is taken as the multi-scale module, which was proposed by he et al. [45]. The advantages of the proposed multi-scale approach can be summarized as follows. The first, small-scale global pooling retains smaller lesions information as much as possible during pooling. The second, based on the three feature maps at different scales, the networks not only pay attention to the local lesions but also focus on the whole breast. Compared with the current method, our method improves the classification performance without using ROI annotations.

3. Methods

The networks structure of the proposed method is shown in Fig. 3. It consists of the following four modules.

(1) The original mammogram images are preprocessed by the BRS module. The module aims to crop out the background regions that are irrelevant for classification and improve the proportion of effective pixels in the image.

(2) Pre-trained CNN is constructed to extract mammogram features. DenseNet [24] and MobileNet [25] are taken as feature extraction modules, respectively,.

(3) To improve the diagnostic process of CNN, a multi-scale module is proposed to replace the last pooling layer of the CNN. It generates three different scales of feature maps, which make CNN can focus on the information of the whole breast and the suspected lesions for classifying.

(4) A fully connected layer is used as the classifier.

The data used in this paper contains mammogram images and corresponding category labels $M = (X, Y)$, where X is the image space and Y represents a two-class space. The mammogram dataset has N samples, and for each sample in the dataset M, the ith image is defined as $x_i (i \in (1, 2 \ldots, N))$. Its corresponding label is y_i.

The mammogram classification task can be formally denoted as: $F : X \rightarrow P$. The model takes the MLO or CC images as input, and the predicted category of the image is the output. P is the predicted category space. For the i-th image x_i in the dataset M, its corresponding predicted category is $p_i (i \in (1, 2 \ldots, N))$. All images will fall into two categories: those with malignant masses and those without any masses. The images with malignant masses are defined as positive samples and the rest are defined as negative samples.

Specifically, the formula can be defined in detail as follows:

$$F : f_C(f_{MF}(f_E(f_{BRS}(X)))) \rightarrow P, \tag{1}$$

where f_C, f_{MF}, f_E and f_{BRS} are the classifier module, the multi-scale module, the feature extraction module and the BRS module, respectively. Among them, the DenseNet f_{DE} and the MobileNet f_{ME} are taken as the feature extraction module, respectively.

First, regardless of whether the image is an MLO view or CC view, the background of mammogram image is relatively large. As shown in Fig. 4, the BRS module consists of three steps, including breast edge detection, breast region cropping, and resizing the image to a standard size, such as 600 × 600 or 224 × 224 pixels. The BRS module can be formally denoted as: $f_{BRS} : X \rightarrow X^*$, where X is original image space and X^* represents the processed image space.

Compared to X, X^* removes most of the redundant information in the original images, which makes the feature extraction

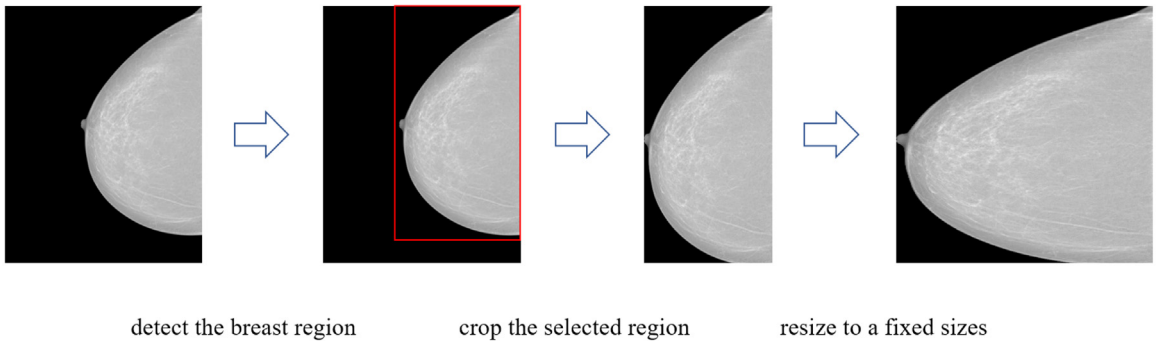


Fig. 4. The BRS module.

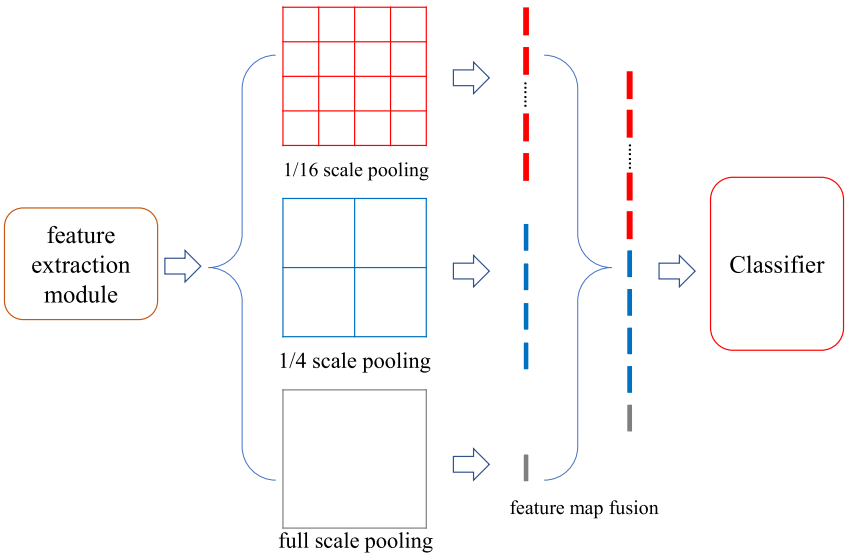


Fig. 5. The multi-scale module.

networks focus on the breast region. The BRS module improves the proportion of effective pixels of the input images.

Then, a feature extraction module is built. A pre-trained DenseNet121 was used to extract features, but only the convolutional and pooling layers before its last pooling layers were used. These layers are fine-tuned to make them focus better on the characteristics of mammogram images. The feature extractor is denoted as f_{DE}, and for each image x^* in X^*, $f_{DE} : x^* \rightarrow x_{fm}$, x_{fm} represents the feature maps generated by the feature extraction module.

Moreover, considering that DenseNet consumes a lot of computing resources, which are scarce in some primary hospitals. It is replaced by MobileNet, which only uses the convolutional layers before its pooling layer as the feature extraction module. The feature extractor of MobileNet is defined as $f_{ME}, f_{ME} : x^* \rightarrow x_{fm}$.

The structure of the multi-scale module is shown in Fig. 5. It is added following the feature extraction module. In the multi-scale module, the feature maps perform max-pooling on three scales: the full image, a quarter image, and a one-sixteenth image. Correspondingly, the three feature maps are generated from three different scales. Different feature map scales correspond to different views. Compared with the original pooling layer, which outputs feature maps at one scale, the multi-scale module obtains 16 small, four middle and one large feature maps. The large one has a global view that corresponds with the doctor's observation of the whole mammogram image. The smaller feature maps focus on small views, corresponding with the doctor's observations of local suspected lesions, and avoids the problem of losing little lesions information during the pooling process. The mesoscale feature map connects these two scales, corresponding with the doctor looking for suspected lesions in the breast. Hence, the proposed CNN has both a global view and local lesions information when classifying images, which improves model performance. The multi-scale module is denoted as f_{MF}, for each feature map x_{fm}, $f_{MF} : x_{fm} \rightarrow x_{mfm}$, x_{mfm} represents the feature maps generated by the feature extraction module.

Finally, the classifier is constructed of a fully connected layer and the sigmoid activation function, which can be formally denoted as: $f_C : x_{mfm} \rightarrow p$, where x_{mfm} is the feature vector generated by the multi-scale module, and p is the predicted category corresponding to this image.

The loss function of the proposed model is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log p_i + (1 - y_i) \cdot \log\left(1 - p_i\right)\right], \tag{2}$$

where $y_i \in \{0, 1\}$ is the label of input x_i(1 is the label of a positive sample), and p_i is the prediction of input x_i.

The pseudocode of the multi-scale neural networks algorithm is given in Algorithm 1.

Algorithm 1 Framework of multi-scale neural networks model.

Input: The raw mammogram images $x_i(i \in (1, 2 \ldots, N))$
Output: The prediction of input images $x_i(i \in (1, 2 \ldots, N))$

```
ending epochs = 200
Initialize the model using pre-trained parameters
while training epochs < ending epochs do
    for a mammogram images x in dataset M do
        Preprocessing images, f_BRS : x → x*
        Extracting features via CNN, f_DE : x* → x_fm
        Generating multi-scales features, f_MF : x_fm → x_mfm
        Classification, f_C : x_mfm → p
        Updating gradients with back propagation algorithm
    end for
end while
while training epochs = ending epochs do
    Save the model and parameters
end while
```

4. Experiments

4.1. Results

The proposed method is evaluated on the public dataset INbreast [18,46], which contains multiple view images of 115 cases for a total of 410 images. Of these, 116 images containing benign or malignant masses are defined as positive samples. While the rest of them are defined as negative samples. Compared to other mammogram datasets, such as the mini-MIAS [47] and DDSM [48] dataset, INbreast's images are original images with higher quality. The images in the dataset are full-field digital mammograms with precise annotations and include several types of lesions (masses, calcifications, asymmetries, and distortions). However, this study uses only the original image and category labels (without ROI annotations).

Table 1 lists the training details. To lowering overfitting, all mammogram images are randomly rotated from −30 degrees to +30 degrees during the training. Because the CNN models have been trained on ImageNet, this study uses the Adam update rules

Table 1
Training details.

Rotated angle	−30∼30
Learning rate	4×10^{-5}
Decay rate	0.98
Batch size	16
Epochs	200
Input sizes	600 × 600 and 224 × 224

and initializes a tiny learning rate of 4×10^{-5}. The learning rate decays at a rate of 0.98 every five epochs. The model will stop training at the 200th epoch. The performance of the proposed methods is evaluated by five-fold cross-validation with classification accuracy (ACC) and the area under the curve (AUC). Besides, the recall (also called the true positive rate) and precision are calculated [49]. The criteria are defined as follows:

$$\begin{aligned} ACC &= \frac{TP + TN}{TP + TN + FP + FN} \\ recall &= \frac{TP}{TP + FN} \\ precision &= \frac{TP}{TP + FP} \end{aligned} \tag{3}$$

where TP, TN, FP, FN are the number of true positives, true negatives, false positives and false negatives, respectively.

Statistical analysis of the accuracy of the model is performed, and the statistical significance level is set to α = 0.05. There are significant differences between the models if p < 0.05, and no significant differences if p > 0.05.

Table 2
Compared with the state of the art models.

Method	ACC(%)	AUC	p-value	ROI
Proposed DenseNet+MS+BRS	96.34	0.9713	0.015	N
Proposed MobileNet+MS+BRS	95.12	0.9653	0.029	N
Sparse MIL [21]	90	0.8586	N/A	N
FCN [22]	**	0.96	N/A	Y
RF+CNN [23]	95	0.91	N/A	Y

Table 2 compares the performance of the proposed method with those of some state-of-the-art methods. The results show that our model achieves the best performance. Sparse MIL [21] is a model without ROI annotations-assisted training. Although it is an automated model, but it does not perform well. RF+CNN [23] and FCN [22] use the ROI annotations-assisted training and perform well. However, they are semi-automatic or manual models that are difficult to be applied in clinical practice. The *p*-value indicates the difference between the proposed methods and the sparse MIL model. The results show that there are significant differences between the proposed models and the sparse MIL model.

The proposed methods with DenseNet achieve a recall of 95.24% and a precision of 90.91% on a test set. The proposed methods with MobileNet achieve a recall of 95.24% and a precision of 86.97% on a test set. Both of the proposed models achieved high recall, which means that our models are very good at detecting positive samples. This is clinically meaningful, as few positive samples are missed diagnoses. MobileNet has a lower precision, which means that it is easier to classify negative samples into positive ones. Its poor feature extraction ability leads to instability performance.

For the reason that the proposed model is based on the combination of global information and local information, rather than just focusing on local features for diagnose. The multi-scale model without ROI annotations-assisted training performs better than the other state-of-the-art models.

4.2. Ablation experiments

Next, the results of several ablation experiments are present.

Table 3
Performance of the proposed multi-scale module.

Image size	224 × 224		600 × 600	
Performance	ACC(%)	AUC	ACC(%)	AUC
MobileNet	84.14	0.8250	92.68	0.8774
MobileNet+MS	91.46	0.9218	95.12	0.9460
DenseNet	86.59	0.8667	90.24	0.9185
DenseNet+MS	93.90	0.9355	95.12	0.9677

The first ablation experiment verifies the performance of the multi-scale module (MS) and the effect of the size of input image on the model. Table 3 shows that the performances of the pre-trained DenseNet and MobileNet on INbreast with 224 × 224 images as input are not very good. However, if the multi-scale modules are incorporated into the models, both the ACC and AUC are substantially improved. DenseNet with the multi-scale module and 224 × 224 input yields an ACC of 93.90% and AUC of 0.9355. These results imply that the multi-scale module produces multi-scale feature maps that help the classifier to merge a global view with local information. This is a very effective way to improve mammogram classification performance.

In the four groups of comparison experiments, model performances with 600 × 600 images are better than those with 224 × 224 images. Even if DenseNet and MobileNet are used without any enhancements, they also perform well. These results indicate that the quality of the mammogram images has a certain impact on the performance. In this group of experiments, the best performance model is obtained by the DenseNet with the multi-scale module on 600 × 600 images (ACC = 95.12% and AUC = 0.9677).

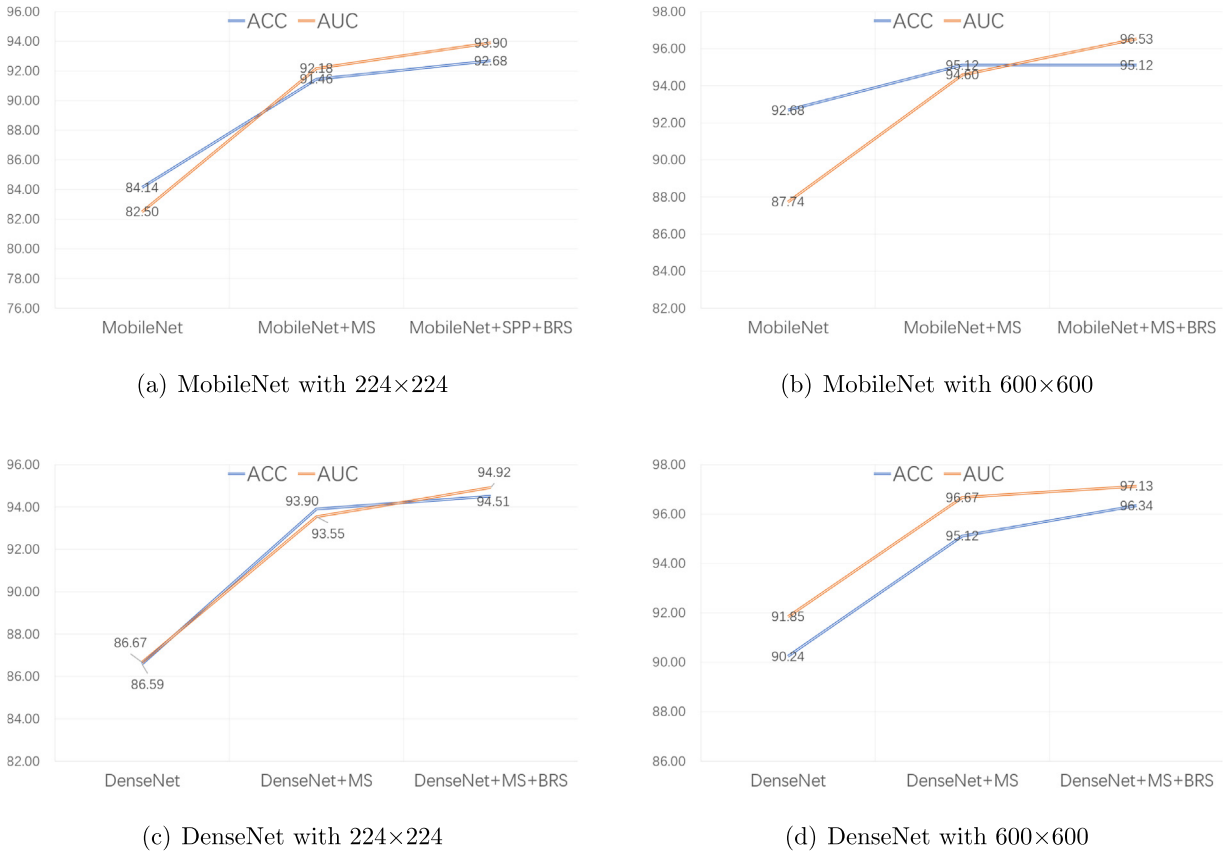
The next verifies the effectiveness of the proposed BRS module. This part of the experiments chose DenseNet and MobileNet with the multi-scale module as the base models. The test samples are images processed by the BRS module.

Table 4
Performance of the proposed BRS module.

Image size	224 × 224		600 × 600	
Performance	ACC(%)	AUC	ACC(%)	AUC
MobileNet+MS	91.46	0.9218	95.12	0.9460
MobileNet+MS+BRS	92.68	0.9390	95.12	0.9653
DenseNet+MS	93.90	0.9355	95.12	0.9677
DenseNet+MS+BRS	94.51	0.9492	96.34	0.9713

As one can see from Table 4, the performance of the model is slightly improved by using the BRS module. The ACC is 96.34% and AUC is 0.9713 on 600 × 600 images. If the size of the image input is 224 × 224 pixels, the performance is slightly lower (ACC = 92.68% and AUC = 0.9355). These results show that the redundant regions of the background hurt the model's mammogram classification performance, and the proposed method is very helpful to solve it.

Fig. 6 shows the performance improvement of the multi-scale module and BRS module. Figs. 6(a) and 6(b) show the performance improvement of MobileNet, and Figs. 6(c) and 6(d) show the DenseNet's improvement. Regardless of whether the networks are DenseNet or MobileNet or whether the input is 224 × 224 or 600 × 600, the multi-scale and BRS modules are effective for mammogram classification. The multi-scale module has a substantial performance improvement on the four models, whereas the BRS module has a certain amount of improvement based on the multi-scale module.



(a) MobileNet with 224×224

(b) MobileNet with 600×600

(c) DenseNet with 224×224

(d) DenseNet with 600×600

Fig. 6. Model performance comparison.

Table 5
Model performance comparison.

Image size	600 × 600	
Model	DenseNet+MS+BRS	MobileNet+MS+BRS
Number of parameters	7,038,529	2,259,165
Time of one epoch	62S	26S
ACC(%)	96.34	95.12
AUC	0.9713	0.9653

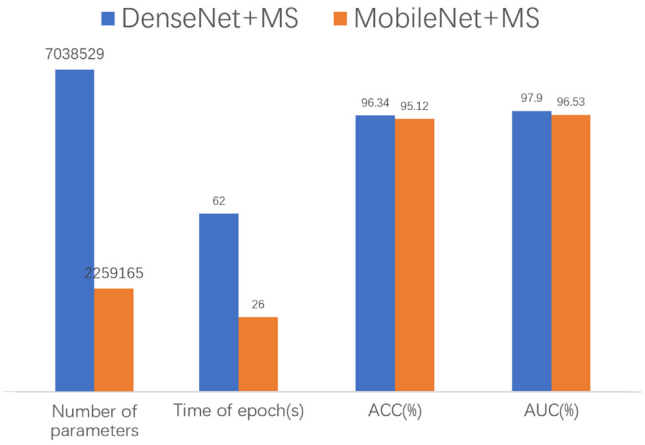


Fig. 7. Performance comparison between MobileNet and DenseNet combined with MS and BRS modules.

Table 5 and Fig. 7 compare the performance of DenseNet and MobileNet along with the multi-scale and BRS modules. The MobileNet's ACC and AUC are slightly lower than those of DenseNet, but MobileNet's parameters and computing time are substantially lower. This experiment aims to explore how to achieve good model performance under the condition of limited computing resources. MobileNet can save 60% of the parameters and computing time with only a slight drop in performance. It may hence be a good choice for hospitals with limited computing resources.

The effectiveness of the proposed method is verified according to the above experiments. To begin with, it is very effective to simulate the doctor's diagnosis by the multi-scale module, which greatly improves the classification performance. Secondly, the original images processed by the BRS module make the model can better focus on the breast region and further improve the performance. Finally, MobileNet may has a slightly weaker feature extraction capability than DenseNet, but it still achieves excellent performance when combined with the multi-scale and BRS modules. Our model has shown to achieve state-of-the-art classification performance on the INbreast dataset.

5. Conclusion

Based on the analysis of the doctor's diagnostic process and the characteristics of mammogram images, an automated multi-scale CNN model for mammogram classification is proposed in this paper. The diagnosis requires to combine features of global breast and suspected lesions. Hence, the multi-scale method is proposed to generates the multi-scale feature maps, which makes CNN not only pay attention to the features of the local lesions, but also combine with the comparison to the global features of breasts. The BRS module is proposed to preprocess the raw mammogram images based on the characteristics of mammogram

images. The processed images contain fewer non-breast pixels and the size of them is small, which contributes to the model to focus on the breast region. The proposed method achieves excellent classification performance and gets the state-of-the-art classification performance on the INbreast dataset. Moreover, the multi-scale approach is applied to the networks with fewer parameters not only can achieve comparable performance but also save 60% of the computing resources. In summary, the multi-scale method can work for both performance and computational efficiency.

For future research, it is promising to extend the current work on the following: (1), the multi-scale methods can be applied in multi-view mammogram classification tasks since mammograms are photographed from four views. (2), the multi-scale methods can be applied to the low-quality mammogram images and improve the performance. (3), the ROI annotation can be combined with the multi-scale methods to train a reliable mass detection model. Besides, the multi-scale methods should be generally applicable to the studies that require a combination of global and local features. And our methods are also applicable to large datasets to improve performance.

CRediT authorship contribution statement

Lizhang Xie: Investigation, Formal analysis, Conceptualization, Ideas, Software, Methodology, Writing - original draft. **Lei Zhang:** Funding acquisition, Supervision, Project administration, Writing - review & editing. **Ting Hu:** Investigation, Data curation, Visualization. **Haiying Huang:** Software, Validation. **Zhang Yi:** Writing - review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants No. 61772353.

References

[1] B. Freddie, F. Jacques, S. Isabelle, S.R. L., T.L. A., J. Ahmedin, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: A Cancer Journal for Clinicians (2015).
[2] S.M. Moss, L. Nystrom, H. Jonsson, E. Paci, E. Lynge, S. Njor, M. Broeders, The impact of mammographic screening on breast cancer mortality in europe: A review of trend studies, J. Med. Screen. 19 (suppl 1) (2012) 26.
[3] C. Wanqing, S. Kexin, Z. Rongshou, Z. Hongmei, Z. Siwei, X. Changfa, Y. Zhixun, L. He, Z. Xiaonong, H. Jie, Cancer incidence and mortality in China, 2014, Chin. J. Cancer Res. 30 (1) (2018) 1.
[4] M.L. Giger, N. Karssemeijer, J.A. Schnabel, Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer, Annu. Rev. Biomed. Eng. 15 (1) (2013) 327–357.
[5] E.A. Sickles, Breast cancer screening outcomes in women ages 40-49: Clinical experience with service screening using modern mammography, J. Nat. Cancer Inst. Monogr. 1997 (22) (1997) 99–104.
[6] B. Lauby-Secretan, C. Scoccianti, D. Loomis, L. Benbrahim-Tallaa, V. Bouvard, F. Bianchini, K. Straif, Breast-cancer screeningviewpoint of the IARC working group, N. Engl. J. Med. 372 (24) (2015) 2353–2358.
[7] R. Bellotti, F. De Carlo, S. Tangaro, G. Gargano, G. Maggipinto, M. Castellano, R. Massafra, D. Cascio, F. Fauci, R. Magro, et al., A completely automated CAD system for mass detection in a large mammographic database, Med. Phys. 33 (8) (2006) 3066–3075.
[8] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, M. Roffilli, A novel featureless approach to mass detection in digital mammograms based on support vector machines, Phys. Med. Biol. 49 (6) (2004) 961–975.
[9] G.M. Brake, N. Karssemeijer, J.H. Hendriks, An automatic method to discriminate malignant masses from normal tissue in digital mammograms., Phys. Med. Biol. 45 (10) (2000) 2843–2857.
[10] M.S. Bae, W.K. Moon, J.M. Chang, H.R. Koo, W.H. Kim, N. Cho, A. Yi, B. La Yun, S.H. Lee, M.Y. Kim, et al., Breast cancer detected with screening US: Reasons for nondetection at mammography, Radiology 270 (2) (2014) 369–377.
[11] Z. Jiao, X. Gao, Y. Wang, J. Li, A deep feature based framework for breast masses classification, Neurocomputing 197 (2016) 221–231.
[12] R. Agarwal, O. Diaz, X. Lladó, R. Martí, Mass detection in mammograms using pre-trained deep learning models, in: 14th International Workshop on Breast Imaging (IWBI 2018), volume 10718, International Society for Optics and Photonics, 2018, p. 107181F.
[13] S. Sasikala, M. Bharathi, M. Ezhilarasi, M. Ramasubba Reddy, S. Arunkumar, Fusion of MLO and CC view binary patterns to improve the performance of breast cancer diagnosis, Curr. Med. Imaging Rev. 14 (4) (2018) 651–658.
[14] X. Gu, Z. Shi, J. Ma, Multi-view learning for mammogram analysis: Autodiagnosis models for breast cancer, in: 2018 IEEE International Conference on Smart Internet of Things (SmartIoT), IEEE, 2018, pp. 149–153.
[15] I. Domingues, P.H. Abreu, J. Santos, Bi-rads classification of breast cancer: A new pre-processing pipeline for deep models training, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 1378–1382.
[16] X. Shu, L. Zhang, Z. Wang, Q. Lv, Z. Yi, Deep neural networks with region-based pooling structures for mammographic image classification, IEEE Trans. Med. Imaging (2020).
[17] Z. Wang, L. Zhang, X. Shu, Q. Lv, Z. Yi, An end-to-end mammogram diagnosis: A new multi-instance and multi-scale method based on single-image feature, IEEE Trans. Cogn. Dev. Syst. (2020).
[18] I.C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M.J. Cardoso, J.S. Cardoso, Inbreast: Toward a full-field digital mammographic database, Academic Radiol. 19 (2) (2012) 236–248.
[19] J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Feifei, Imagenet: A large-scale hierarchical image database, 2009, pp. 248–255.
[20] T. Lin, M. Maire, S.J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, C.L. Zitnick, Microsoft COCO: Common objects in context, European Conference on Computer Vision (2014) 740–755.
[21] W. Zhu, Q. Lou, Y.S. Vang, X. Xie, Deep multi-instance networks with sparse label assignment for whole mammogram classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 603–611.
[22] L. Shen, End-to-end training for whole image breast cancer diagnosis using an all convolutional design, 2017, arXiv preprint arXiv:1708.09427.
[23] N. Dhungel, G. Carneiro, A.P. Bradley, The automated learning of deep features for breast mass classification from mammograms, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 106–114.
[24] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[25] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, ArXiv: Computer Vision and Pattern Recognition (2017).
[26] Z. Yi, Foundations of implementing the competitive layer model by Lotka–Volterra recurrent neural networks, IEEE Trans. Neural Netw. 21 (3) (2010) 494–507.
[27] L. Zhang, Z. Yi, Dynamical properties of background neural networks with uniform firing rate and background input, Chaos Solitons Fractals 33 (3) (2007) 979–985.
[28] L. Zhang, Z. Yi, S.-i. Amari, Theoretical study of oscillator neurons in recurrent neural networks, IEEE Trans. Neural Netw. Learn. Syst. 29 (11) (2018) 5242–5248.
[29] B. Zhao, X. Zhang, H. Li, Z. Yang, Intelligent fault diagnosis of rolling bearings based on normalized cnn considering data imbalance and variable working conditions, Knowl.-Based Syst. (2020) 105971.
[30] S. Lee, H. Kim, Q.X. Lieu, J. Lee, CNN-Based image recognition for topology optimization, Knowl.-Based Syst. (2020) 105887.
[31] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[32] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Int. Conf. Learn. Represent. (2015).
[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[35] L. Wang, L. Zhang, Z. Yi, Trajectory predictor by using recurrent neural networks in visual tracking, IEEE Trans. Cybern. 47 (10) (2017) 3172–3183.
[36] Y. Wang, L. Zhang, L. Wang, Z. Wang, Multitask learning for object localization with deep reinforcement learning, IEEE Transactions on Cognitive and Developmental Systems 11 (4) (2018) 573–580.
[37] J. Wang, R. Ju, Y. Chen, L. Zhang, J. Hu, Y. Wu, W. Dong, J. Zhong, Z. Yi, Automated retinopathy of prematurity screening using deep neural networks, EBioMedicine 35 (2018) 361–368.
[38] J. Mo, L. Zhang, Y. Feng, Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks, Neurocomputing 290 (2018) 161–171.
[39] A. Esteva, B. Kuprel, R.A. Novoa, J.M. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.
[40] X. Qi, L. Zhang, Y. Chen, Y. Pi, Y. Chen, Q. Lv, Z. Yi, Automated diagnosis of breast ultrasonography images using deep neural networks, Med. Image Anal. 52 (2019) 185–198.
[41] J. Hu, Y. Chen, J. Zhong, R. Ju, Z. Yi, Automated analysis for retinopathy of prematurity by deep neural networks, IEEE Trans. Med. Imaging 38 (1) (2019) 269–279.
[42] W. Zhu, C. Liu, W. Fan, X. Xie, Deeplung: Deep 3D dual path nets for automated pulmonary nodule detection and classification, Workshop Appl. Comput. Vis. (2018) 673–681.
[43] L. Wang, L. Zhang, M. Zhu, X. Qi, Z. Yi, Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks, Med. Image Anal. 61 (2020) 101665.
[44] H. Li, D. Chen, W.H. Nailon, M.E. Davies, D. Laurenson, Improved breast mass segmentation in mammograms with conditional residual u-net, in: Image Analysis for Moving Organ, Breast, and Thoracic Images, Springer, 2018, pp. 81–89.
[45] 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 (9) (2015) 1904–1916.
[46] A.N. Kia, S. Haratizadeh, S.B. Shouraki, Network-based direction of movement prediction in financial markets, Eng. Appl. Artif. Intell. 88 (2020) 103340.
[47] P. Suckling J., The mammographic image analysis society digital mammogram database, Digital Mammo (1994) 375–386.
[48] C. Rose, D. Turi, A. Williams, K. Wolstencroft, C. Taylor, Web services for the DDSM and digital mammography research, in: International Workshop on Digital Mammography, Springer, 2006, pp. 376–383.
[49] A. Fernandez, S. Garcia, M.J.D. Jesus, F. Herrera, A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets, Fuzzy Sets and Systems 159 (18) (2008) 2378–2398.