



Multiview multimodal network for breast cancer diagnosis in contrast-enhanced spectral mammography images

Jingqi Song¹ · Yuanjie Zheng¹ · Muhammad Zakir Ullah¹ · Junxia Wang¹ · Yanyun Jiang¹ · Chenxi Xu¹ · Zhenxing Zou² · Guocheng Ding²

Received: 28 December 2020 / Accepted: 27 April 2021 / Published online: 8 May 2021
© CARS 2021

Abstract

Purpose CESM (contrast-enhanced spectral mammography) is an efficient tool for detecting breast cancer because of its image characteristics. However, among most deep learning-based methods for breast cancer classification, few models can integrate both its multiview and multimodal features. To effectively utilize the image features of CESM and thus help physicians to improve the accuracy of diagnosis, we propose a multiview multimodal network (*MVMM-Net*).

Methods The experiment is carried out to evaluate the in-house CESM images dataset taken from 95 patients aged 21–74 years with 760 images. The framework consists of three main stages: the input of the model, image feature extraction, and image classification. The first stage is to preprocess the CESM to utilize its multiview and multimodal features effectively. In the feature extraction stage, a deep learning-based network is used to extract CESM images features. The last stage is to integrate different features for classification using the *MVMM-Net* model.

Results According to the experimental results, the proposed method based on the Res2Net50 framework achieves an accuracy of 96.591%, sensitivity of 96.396%, specificity of 96.350%, precision of 96.833%, F1_score of 0.966, and AUC of 0.966 on the test set. Comparative experiments illustrate that the classification performance of the model can be improved by using multiview multimodal features.

Conclusion We proposed a deep learning classification model that combines multiple features of CESM. The results of the experiment indicate that our method is more precise than the state-of-the-art methods and produces accurate results for the classification of CESM images.

Keywords Contrast-enhanced spectral mammography · Classification · Breast cancer · Multiview · Multimodal

Introduction

Breast cancer is one of the most frequent female malignancies, which seriously affects their health [1]. According to Global Cancer Statistics [2] among all the new cases of cancers, 30% will be diagnosed as breast cancer and 15% of deaths will be caused due to breast cancer in 2021. Due to the unclear causes and effective prevention of breast cancer, breast cancer cases have not been significantly reduced. But, diagnosis of breast cancer in early screening can reduce

the mortality rate. Therefore, the establishment of reliable screening methods is essential for the early diagnosis of breast cancer [3].

Contrast-enhanced spectral mammography (CESM, also known as contrast-enhanced digital/dual-energy mammography) is an emerging technique to detect breast cancer [4]. The CESM images are acquired through a digital dual-energy CESM device which is derived from a standard Digital Senographe Essential Mammography [5]. The prototype of CESM device is a full-field digital mammography system that is modified to allow dual-energy exposures. It comprises a flat panel detector with a cesium iodide absorber, dedicated software, and hardware. The standard CESM procedure consists of the following steps: an iodinated contrast agent (300mg iodine/mL, 1.5mL/kg of body weight, flow-rate 3mL/s) is administered to the patient. Two minutes after contrast injection, the patient undergoes a breast examination

✉ Yuanjie Zheng
yjzheng@sdnu.edu.cn

¹ School of Information Science and Engineering, Shandong Normal University, Jinan, China

² Medical Imaging Department, Yantai Yuhuangding Hospital, Yantai, China

with the help of medical staff. A set of craniocaudal (CC) and mediolateral oblique (MLO) views of the breast is then collected. After the injection the analysis is carried out by two automatic X-ray energy levels: low-energy (26–32 kVp) X-ray spectrum and high-energy (45–49 kVp) X-ray spectrum [6]. Then, the low-energy (LE) and high-energy (HE) images are reconstructed by a specific image reconstruction algorithm to obtain the dual-energy subtracted (DES) image that can suppress background texture and highlight contrast uptake. HE images are rarely studied images due to their low clinical value. Therefore, CESM consists of two completely different modalities of image: LE image and DES image. The typical CESM images are shown in Fig. 1. Because of its multimodal characteristic, CESM provides high contrast and lesion-visible image [7], which is more advantageous in detecting breast cancer. [8] and [9] have presented that CESM can improve the diagnostic performance of radiologists in clinical practice. Numerous studies [10–12] have shown that CESM has higher sensitivity and excellent receiver operating characteristic (ROC) curve areas for the diagnosis of breast cancer as compared to mammography. It overcomes the tissue overlap phenomenon and provides tumor information with high image resolution similar to magnetic resonance imaging (MRI), which is a promising imaging technique [13].

In clinical practice, the detection of breast images by experts is usually a manual process, which is time-consuming and costly [14]. In addition, the radiologist may miss the primary lesion due to the overlapping of dense tissue or changes in the shape, size, and borders of the mass [15]. Machine learning (ML)-based image analysis methods can improve the accuracy of breast cancer detection and can be used as a “second option” to improve the performance of diagnosis of breast cancer [16,17]. In [18], Konen et al. proposed an automatic breast lesion classification method using both mammography and ultrasound images. Shaikh et al. [19] proposed learning-based method using privileged information (LUPI)-based CAD framework for the diagnosis of breast cancer disease. There are fewer ML-based CESM image analysis systems because they are relatively new in imaging technology. Mateos et al. [20] analyzed seventeen textural descriptors from gray-level co-occurrence matrix (GLCM) and summarized the potential of CESM images to improve the diagnostic accuracy of breast cancer. Perek et al. [21] compared two analysis approaches, fine-tuned a pretrained network and fully trained convolutional neural network, to classify breast masses as benign or malignant in CESM images. They also used breast imaging reporting and data system (BI-RADS) [22] textual information as input to improve the classification performance of the network. Danala et al. [10] developed a CAD scheme for segmenting tumor regions on LE and DES images to classify breast masses. Gao et al. [13] proposed a shallow-deep convolutional neural network (SD-CNN), where the shallow

CNN derives “virtual” recombinant images from LE images and the deep CNN extracts new features from LE, “virtual” recombinant images to classify cases into benign and tumor. Another study [9] focused on the lesion region of interest (ROI) and extracted 236 textural features from the CESM. Then, the most noteworthy features were classified using the support vector machines (SVM). Recently, Fanizzi et al. [23] proposed a fully automated support system for diagnosis of breast cancer in CESM Images. Then, different techniques were used to extract the ROI features set of low-energy and recombined images. Through the sequential feature selection algorithm, the random forest classifier was trained on the selected significant feature subset. Losurdo et al. [24] extracted and analyzed 55 ROIs and trained a SVM classifier to classify the CESM. In [25], the author revealed the relationship between the statistical features extracted from CESM images and the histological outcome of breast cancer.

These imaging processing techniques indicate that useful quantitative information can be extracted from CESM images to improve diagnosis accuracy. However, most of the existing methods replace the natural input image with medical image without effectively considering the specific characteristics of CESM. For instance, CESM images have multiple views (CC and MLO view) to provide different information about lesions, which can help in improving the performance of classification. In addition, CESM images are characterized by multimodality (LE and DES modes), which has rarely been considered simultaneously in previous studies. Based on the above considerations, we propose a MVMM-Net framework for distinguishing malignant from benign breast lesions in the CESM images. Compared with previous studies, the proposed MVMM-Net can simultaneously extract features from multiview and multimodal of the same breast image. By combining multiple features, the proposed method achieves a significant improvement in classification performance.

Materials and methods

Dataset

We retrospectively collected the CESM performed between July 2018 and April 2020 of 95 patients aged 21–74. We preprocess the data to remove the sensitive information about the patient. Each patient has eight images, four on each side of the left and right breast. Since the condition of the left and right breast may be different in each patient, we treat the images of a patient as two cases (one case for each breast). Our dataset includes 95 patients, of which 58 are malignant and 132 benign cases. Each case has 4 images, with a total of 760 images. The typical size of CESM images in our experiment is 2394*3062, and these images are divided into

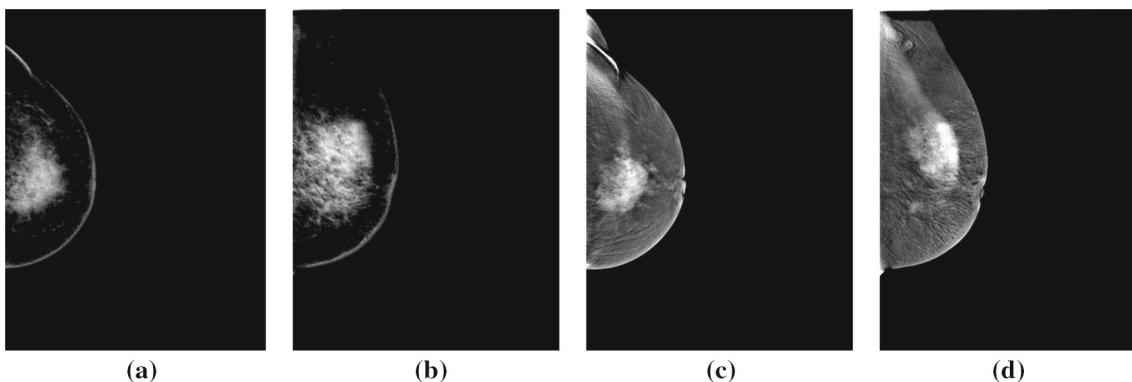
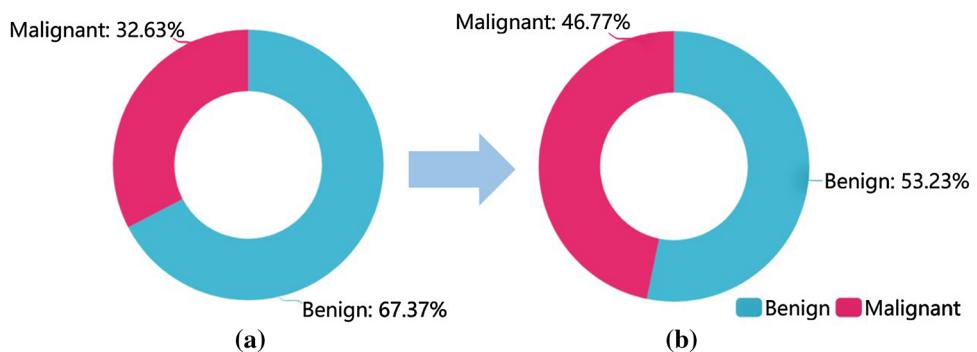


Fig. 1 The typical CESM images of the left breast. **a, b** LE modality image of CC and MLO view. **c, d** DES modality image of CC and MLO view

Fig. 2 The distribution of cases in the experiment. **a** The proportion of benign and malignant cases. **b** The proportion of adjusted benign and malignant cases



two classes (benign/malignant). The proportion of images is shown in Fig. 2a. As can be seen from the figure, the dataset is highly imbalanced. We randomly discard half of the benign image data to produce a more balanced dataset to minimize the potential classification bias towards minority (malignant) cases. By utilizing the random discard strategy, the number of benign cases is narrowed from 132 to 66. The dataset becomes more balanced with 66 (53.2%) benign and 58 (46.8%) malignant cases (Fig. 2b).

CESM images preprocessing

To train sufficiently, preprocessing is a common stage in the neural network. The main goal is to improve the features of the image by applying a set of transformations, which helps to improve the robustness of the model. The specific steps are as follows. The first step is to remove most of the black background. The black background occupies a large part of the image content. However, these backgrounds have no positive impact on our classification. To maximize the removal of the black background while ensuring that the image area of the breast is not removed, we segment the image into the size of 1350*3062. In addition, the classification model always performs well when the dataset is large enough. Then, we randomly select half of the images to flip horizontally or vertically for meeting the model's needs. Besides, we randomly select 20% of the images to add Gaussian noise or rotate them

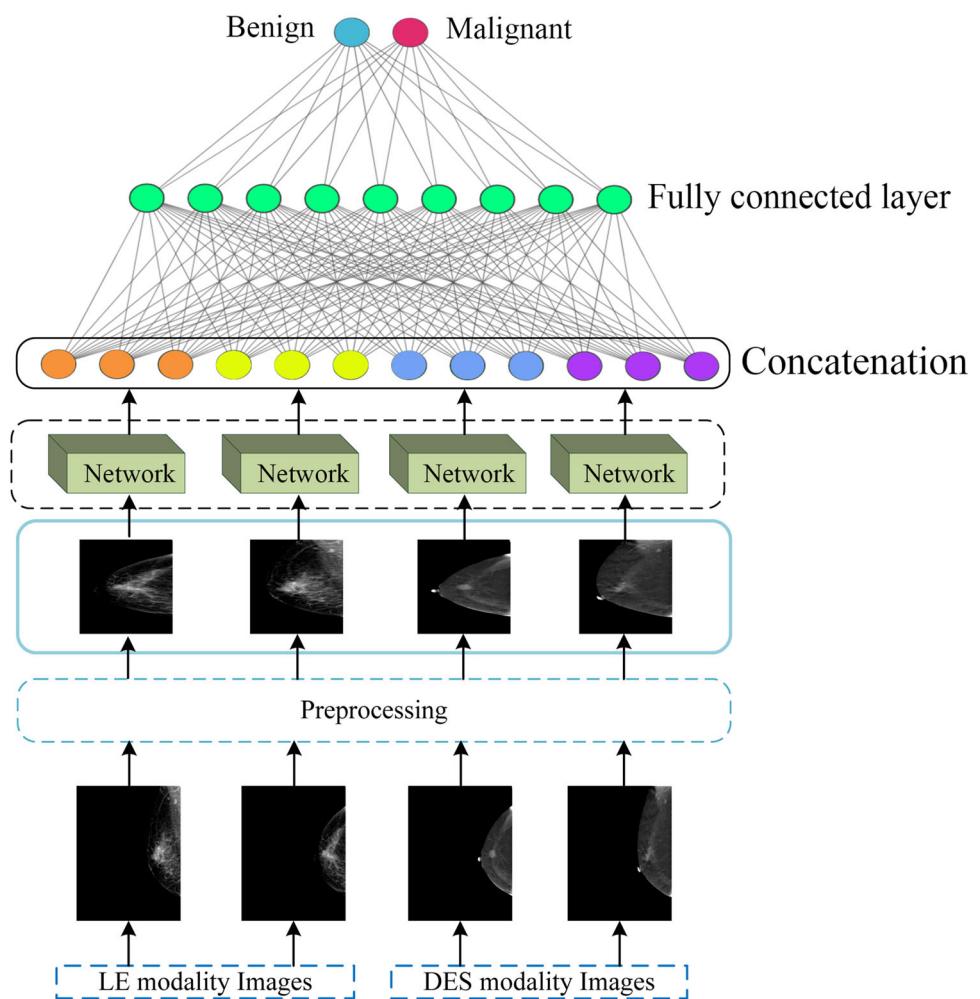
in the range of -40 to 40^{deg} . Due to the different lighting conditions between images during digitization, it affects all pixel values of the image. A common way to overcome this effect is to perform global contrast normalization, where the average of the image intensities is subtracted from each pixel. Local contrast normalization is also performed. Through the above image enhancement methods, the number of the image in our dataset is increased from 496 to 17360. We randomly select 80% of the image for training, 15% for validation, and 5% for final testing. After augmenting, we resize the whole image to a uniform size before feeding it into the network.

The structure of MVMM-Net

As mentioned above, the discrimination performance can be improved by using multiview and multimodal CESM images. Inspired by this motivation, we propose a novel network to classify the CESM images.

We propose a new multiview multimodal network (*MVMM-Net*), which can extract the features of multiview and multimodal images at the same time. As shown in Fig. 3, the model uses four images as inputs, including two views and two modalities. The network can be one of the well-known deep learning architectures, such as VGGNet [26], ResNet [27], SE-ResNet [28], ResNeXt [29], WRN [30], and Res2Net [31]. In our experiments, we find that the 'Res2Net50' model have the highest accuracy on the test set

Fig. 3 Main flowchart of the proposed MVMM-Net



for the benign/malignant classification task (shown in Table 2). Unless we explicitly specify, we will report the results of this model. In the MVMM-Net, the model shares the same network, while holding independent parameters. In detail, the network consisting of a cascaded convolutional layers, batch normalization layers, nonlinear activation layers, and pooling layers. This feed-forward framework can be regarded as a feature extractor. Therefore, multiview and multimodal features can be extracted from the input images simultaneously. Finally, each branch feeds into its respective global average pooling layer to calculate the average of each feature map. Then, all branches of the feature map are concatenated and input to the fully connected layer. A softmax classifier is used to make the final decision of the classification task. We denote the training dataset of our MVMM-Net by $\{(x_i, y_i), i = 1, 2, \dots, N\}$, where x_i denotes the i th set of input CESM image, $y_i \in \{0, 1\}$ denotes the ground truth label (proved by biopsy) assigned to i th input set. We denote the output of classification by $\tilde{y}_i \in \{0, 1\}$. The loss function of classification is defined as

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N H(\tilde{y}_i, y_i) \quad (1)$$

where $H()$ is the cross-entropy function for measuring the dissimilarity between \tilde{y}_i and y_i .

Model setting

In our experiment, the image is resized to 227 * 227 pixels. We set the learning rate as 0.01, and the number of epochs for the training process is set to 100. In the training process, the batch size is set as 16. All methods are optimized by Adam optimizer [32]. The framework of all methods for evaluating is Pytorch in this paper. The software environment is Ubuntu 16.04 and Pytorch 3.7. The hardware environment is Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz and NVIDIA GV100GL Tesla V100 32GB GPU.

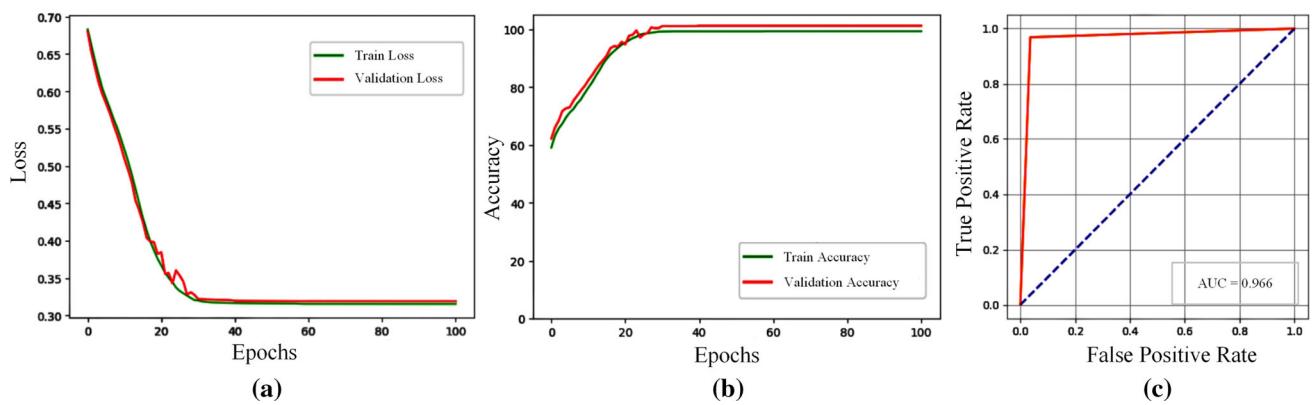


Fig. 4 The loss and accuracy curves on training and validation process and the ROC curve in the test phase. **a** The loss curves. **b** The accuracy curves. **c** The ROC curve

Performance evaluation

To evaluate the performance of the proposed method in classifying CESM images, different evaluation criteria have been used such as accuracy, precision, specificity, sensitivity, and F1_score.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (4)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (5)$$

$$F1_score = \frac{2 * \text{TP}}{\text{FP} + \text{FN} + 2\text{TP}} \quad (6)$$

where TP, FP, TN, and FN are the true positive, false positive, true negative, and false negative rates, respectively. From the aforementioned evaluation metrics, it is stated that, the higher these values are, the better the classification performance of the system. Among these metrics, sensitivity is particularly important. And it is defined as the proportion of samples that are predicted to be positive cases out of all samples that are positive cases. In the medical field, high-risk categories such as diseases are often used as positive categories, and the cost of missing a positive category is very high. In addition to the above five criteria, the ROC curve and AUC are also used to evaluate the performance. The ROC curve near the upper left corner represents the most accurate classifier. AUC is the area under the ROC curve, which is a performance index to measure the quality of the classifier. It indicates the probability that the positive case of prediction is in front of the negative case.

Results

Performance of the proposed model

The loss and accuracy curves of MVMM-Res2Net50 during the training and validation process are presented in Fig. 4a, b. As the training proceeds, the accuracy of classification gradually increases and then reaches the maximum value. At the same time, the loss curve gradually decreases until it tends to be smooth. The training accuracy increases to almost 100% and the loss no longer changes, which indicates the success of our approach. This shows that our method has achieved good results in the training process. Figure 4c is the ROC curve of our model on the test set. The accuracy, sensitivity, specificity, and AUC of MVMM-Res2Net50 reached 96.6%, 96.4%, 96.4%, and 0.966, respectively (shown in Table 1), which indicates the good classification performance of our model.

Class activation mapping

To understand the discriminant region of CESM classification, we visualize the discriminative region using the Grad-Class Activation Mapping (Grad-CAM) [33]. The area covered with lighter color means stronger CAM. Figure 5 is the CAM of CESM images that classified as malignant (first row) and benign (second row) by the proposed method. The highlights in the first row focus on the tumor area, indicating that the model can accurately distinguish the tumor area from the normal area. The lighter areas of normal images are mainly in the whole breast region. As can be seen from the figure, the proposed method identifies the regions that play an important role in the classification. Based on the radiologist's opinion, the location of the CAM thermogram of the lesion obtained by the model also matches the physician's region of interest. The accurate localization of tumor and

Table 1 A comparison of state-of-the-art research and our proposed methodology

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Patel et al. [9]	90.0	88.0	92.0	0.950
Gao et al. [13]	90.0	83.0	94.0	0.920
Perek et al. [21]	—	100.0	66.0	0.890
Danala et al. [10]	78.4	—	—	0.848
Fanizzi et al. [23]	87.5	87.5	91.7	0.931
MVMM-Res2Net50 (ours)	96.6	96.4	96.4	0.966

Bold indicates the results that surpass all competing methods

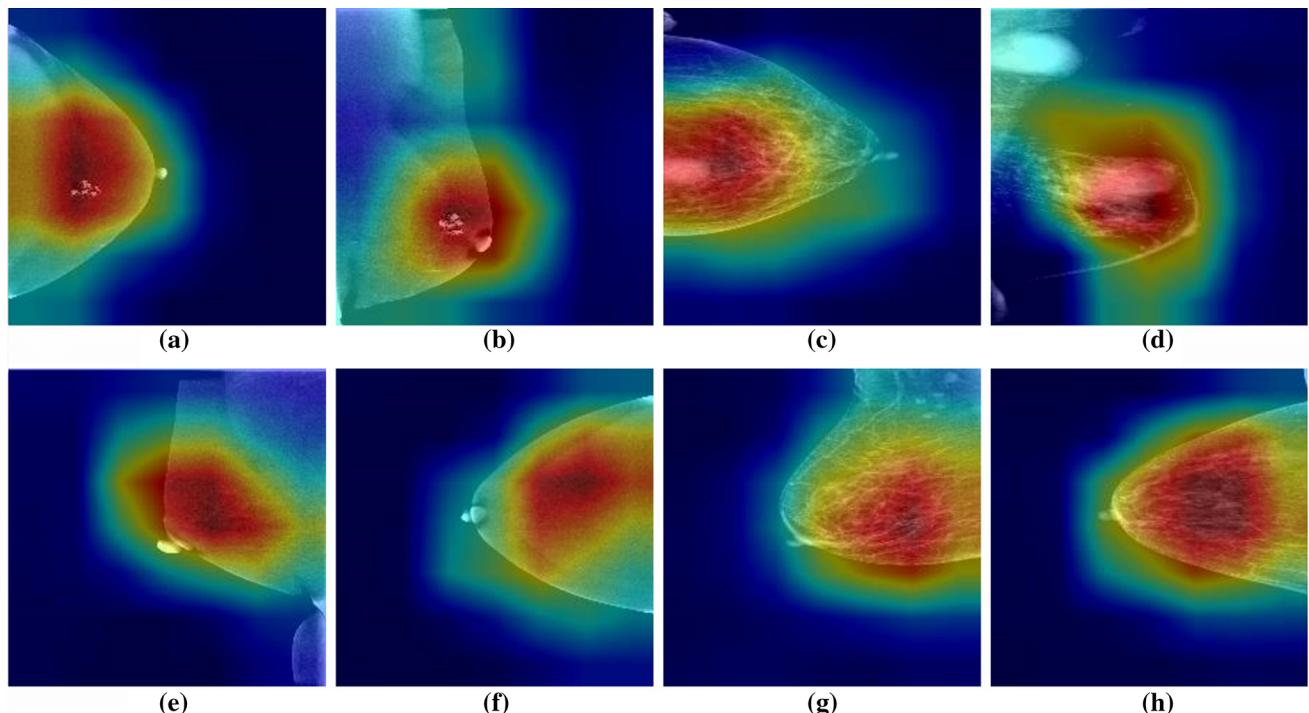


Fig. 5 Visualization of class activation mapping for CESM images. **a–d** The class activation mapping on the cancer breast. **e–h** The class activation mapping on the normal breast

breast regions demonstrates the potential value of our network in CESM image classification.

Performance of the proposed method compared with existing methods

In this section, we compare the MVMM-Res2Net50 network with the state-of-the-art methods. Detailed comparisons are given in Table 1. We compare these methods in terms of four evaluation metrics. Some evaluation indicators not given in the original paper are empty in the table. It can be seen from the table that our evaluation indexes are above 96.4%. Compared with other methods, our approach outperforms almost all other state-of-the-art models. Although the sensitivity of Perek [21] is 100%, the specificity index of this method is only 66%, which is the lowest among all methods. Presum-

ably the method in [21] trained on a relatively small dataset. We hypothesize that the superior performance of the proposed method is related to its ability to efficiently integrate multiview and multimodal features. In Sect. 4, we empirically investigate this hypothesis through multiple ablation studies.

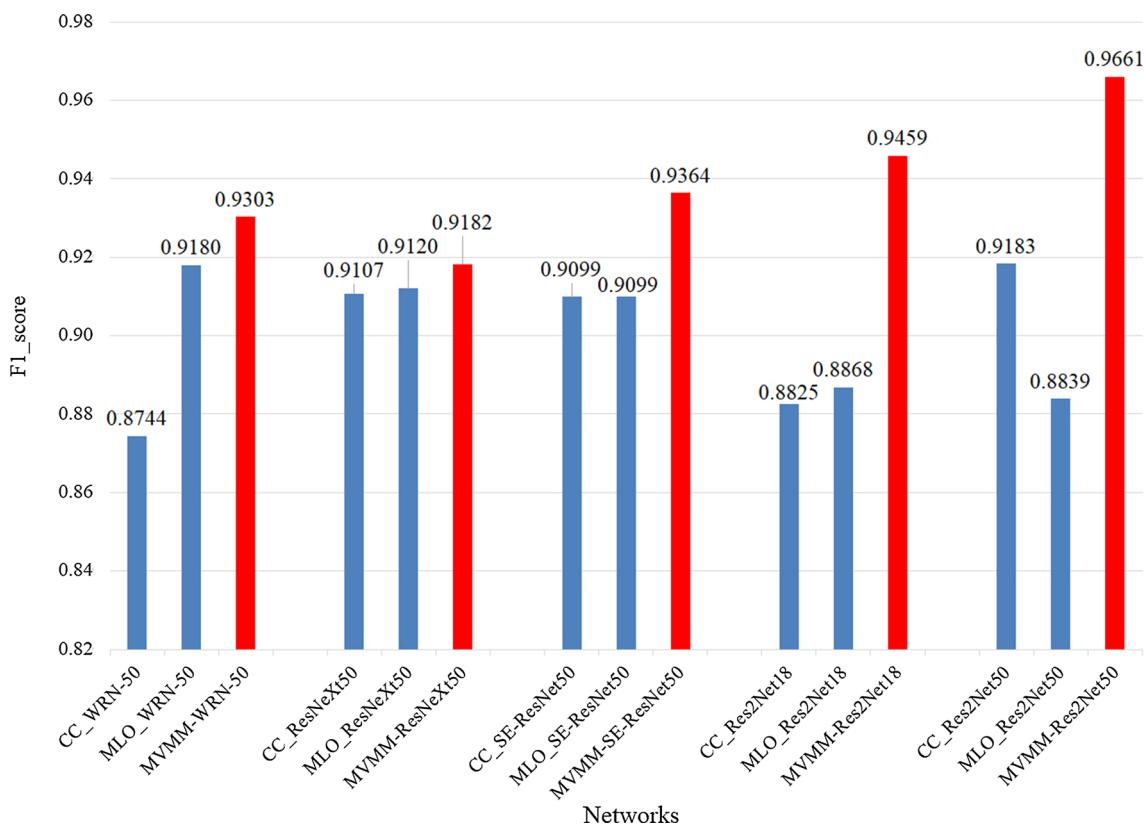
Discussion

In this paper, we present a method that serves as a supporting tool for physicians to advise on the diagnosis of breast cancer and reduce unnecessary biopsies. The experimental results demonstrate that our method can solve the problem of benign/malignant classification of CESM images. In this section, the effects of different views and modalities on the experimental results are discussed.

Table 2 Performance analysis of classification methods under different networks

CNN Architecture	Precision (%)	Specificity (%)	Sensitivity (%)	AUC
MVMM-VGGNet19	82.74	83.64	84.23	0.832
MVMM-ResNet18	95.11	96.28	96.37	0.957
MVMM-ResNet50	93.95	91.11	90.99	0.925
MVMM-SE-ResNet50	94.50	92.66	92.79	0.936
MVMM-ResNeXt50	92.66	90.99	90.99	0.918
MVMM-WRN-50	92.83	93.09	93.24	0.930
MVMM-Res2Net18	94.60	94.50	94.60	0.946
MVMM-Res2Net50	96.83	96.35	96.40	0.966

Bold indicates the results that exceed all competing methods in this column

**Fig. 6** The F1_score of different views in various networks

The performance of MVMM-Net with different networks

In this experiment, we first investigate the effect of different networks on the CESM image classification problem using the test dataset. We modify the classical classification networks according to their article description and compare them with our method. Specifically, we change the input layers of the network for using the multiview and multimodal CESM images. From the statistical analysis shown in Table 2, we summarize the performance of different networks according to their precision, specificity, sensitivity, and AUC.

Overall, the improved classification network achieved high results, with almost all metrics above 90%. The precision, specificity, sensitivity, and AUC of MVMM-Res2Net50 are the highest among the mentioned networks. From the above experiments, we confirm that our proposed structure is suitable for CESM image classification.

The performance of classification under different CESM views

In this part, we design an experiment to evaluate whether the multiview network can improve the performance com-

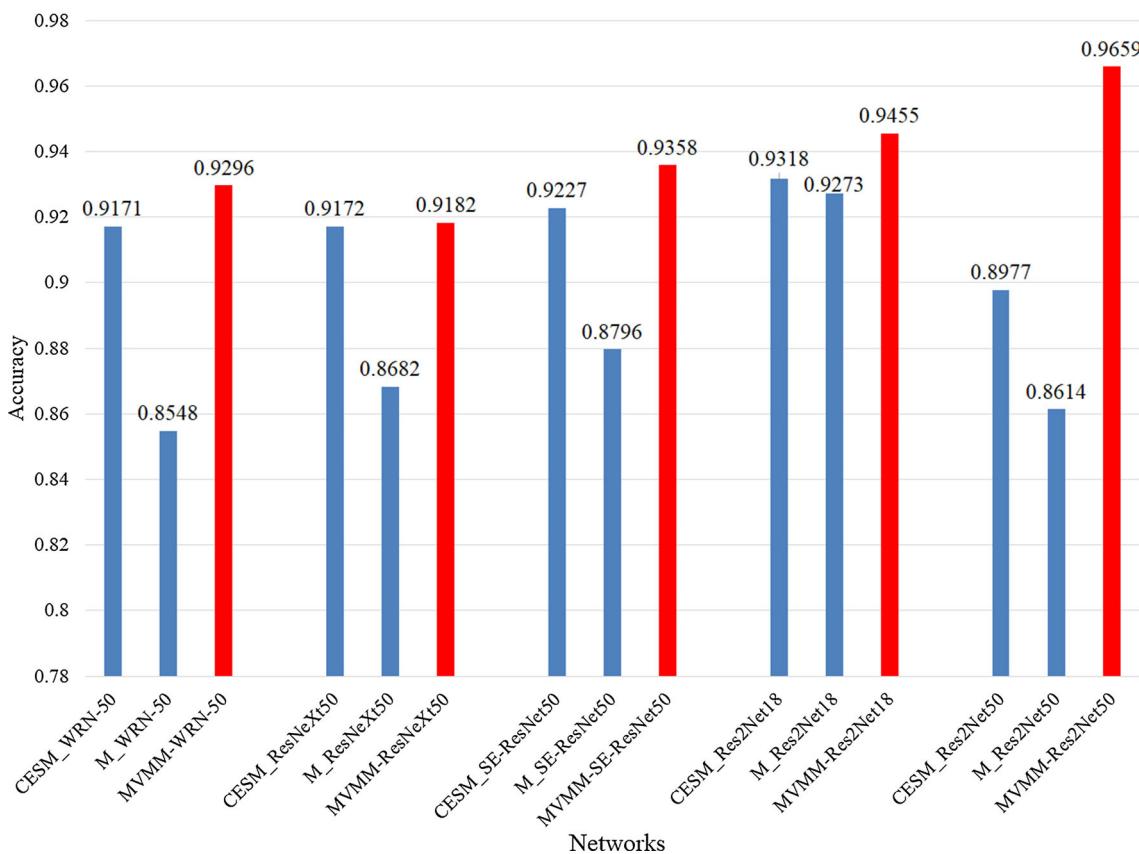


Fig. 7 The accuracy of different modalities in various networks

pared with the single view network. We group the dataset of the same view before doing the comparison experiment. The dataset is divided into two groups based on the view. For single-view classification, only images from the same view (CC or MLO) are fed into the network for training and testing. For multiview classification, we input both views into the network for analysis. We compare the F1_score of these two types of networks. The results are shown in Fig. 6. As can be seen from the figure, the results of the F1_score with multiple views are better than those of the single view, regardless of the type of network. It can be seen that a multiview-based network can improve the performance of CESM image classification.

The comparison of different CESM image modalities on classification performance

In this section, we evaluate the performance of the network for multimodal images versus using a unimodal network. Similar to evaluating the performance of the viewpoints, we also group the datasets based on the modality of the images before performing the evaluation. We train the network using a single modal image (LE or DES image) separately. On the other hand, we also use images from both modalities

simultaneously for feature extraction and analysis. The final evaluation results obtained are shown in Fig. 7. As it is seen from the figure, the classification accuracy of the multimodal network is higher than that of the single modal network. In other words, the network performance is indeed improved when multimodal images are utilized. In addition, the network using DES images for classification obtained significantly better results than the network using LE images, proving that DES images do help to improve the accuracy of diagnosis. Our method proves that DES image is not only helpful to clinical diagnosis, but also can improve the accuracy of CESM images classification.

Conclusions

In this study, we propose a multiview multimodal network to diagnose breast cancer in CESM images. Our proposed method allows the classification of CESM images of the breast to distinguish between benign and malignant cases. With this method, we can obtain lesion information from different views and modalities simultaneously, making the classification results more accurate. Our experimental results have demonstrated that it is essential to classify CESM

images using images in multiple views. We also show experimentally that using CESM images of multiple modalities simultaneously can also improve the classification performance of the network. Our algorithm can provide physicians with more recommendations when diagnosing breast cancer from CESM images and reduce the number of surgeries for benign breast nodules. However, we just take the image as the input of classification, without considering the actual situation of the patient, such as the patient's age and family history. In future work, we need to expand our work by collecting more hospital data and incorporating this information.

Acknowledgements This research was supported by Projects funded by National Natural Science Foundation of China Grant Numbers 81871508 and 61773246, the Major Program of Shandong Province Natural Science Foundation Grant Number ZR2018ZB0419, and the Taishan Scholar Program of Shandong Province of China grant number TSHW201502038.

Declarations

Conflicts of Interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Farzad R, Gordon F, Sahar T, Mahnaz N, Amir A, Soodabeh S (2020) Role of regulatory miRNAs of the pi3K/AKT signaling pathway in the pathogenesis of breast cancer. *Gene* 737(144459):02
- Siegel R, Miller K, Fuchs H, Jemal A (2021) Cancer statistics. *CA Cancer J Clin* 71(1):7–33
- Hiba C, Hamid Z, Omar A (2020) Multi-label transfer learning for the early diagnosis of breast cancer. *Neurocomputing* 392:168–180
- Martin D, Tobias DZ, Wolfram S, Birgit A, Florian K, Werner J, Clarisse D, Willi O, Michael H, Christian M (2015) Dual-energy contrast-enhanced spectral mammography (CESM). *Arch Gynecol Obstet* 292(4):739–747
- Lisa H, Martin D, Christoph J, Elena E, Julia H (2019) Contrast-enhanced spectral mammography with a compact synchrotron source. *PLoS ONE* 14(10):
- James JJ, Tennant SL (2018) Contrast-enhanced spectral mammography (CESM). *Clin Radiol* 73(8):715–723
- Lobbes MBI, Smidt ML, Houwers J, Tjan-Heijnen VC, Wildberger JE (2013) Contrast enhanced mammography: techniques, current results, and potential indications. *Clin Radiol* 68(9):935–944
- Cheung YC, Tsai HP, Lo YF, Ueng SH, Huang PC, Chen SC (2016) Clinical utility of dual-energy contrast-enhanced spectral mammography for breast microcalcifications without associated mass: a preliminary analysis. *Eur Radiol* 26(4):1082–1089
- Patel BK, Ranjbar S, Wu T, Pockaj BA, Li J, Zhang N, Lobbes M, Zhang B, Mitchell JR (2018) Computer-aided diagnosis of contrast-enhanced spectral mammography: a feasibility study. *Eur J Radiol* 98(3):207–213
- Gopichand D, Bhavika P, Faranak A, Morteza H, Jing L, Teresa W, Bin Z (2018) Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms. *Ann Biomed Eng* 46(9):1419–1431
- Dromain C, Thibault F, Diekmann F, Fallenberg EM, Jong RA, Koomen M, Hendrick RE, Tardivon A, Toledano A (2012) Dual-energy contrast-enhanced digital mammography: initial clinical results of a multireader, multicase study. *Breast Cancer Res* 14(3):R94
- Tagliafico AS, Bignotti B, Rossi F, Signori A, Sormani MP, Vadola F, Calabrese M, Houssami N (2016) Diagnostic performance of contrast-enhanced spectral mammography: systematic review and meta-analysis. *Breast* 28:13–19
- Fei G, Teresa W, Jing L, Bin Z, Lingxiang R, Desheng S, Bhavika P (2018) SD-CNN: a shallow-deep CNN for improved breast cancer diagnosis. *Comput Med Imaging Graph* 70:53–62
- Asmaa I, Paul G, Ronnachai J, Abdelsamea Mohammed M, Mermel Craig H, Po HCC, Rakha Emad A (2020) Artificial intelligence in digital breast pathology: techniques and applications. *Breast* 49:267–273
- Arnau O, Jordi F, Joan M, Elsa P, Josep P, Denton Erika RE, Reyer Z (2010) A review of automatic mass detection and segmentation in mammographic images. *Med Image Anal* 14(2):87–110
- Stephanie R, Hossein A, Kevin S, Johan H (2018) Digital image analysis in breast pathology from image processing techniques to artificial intelligence. *Transl Res* 194:19–35
- Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ (2019) Artificial intelligence in breast imaging. *Clin Radiol* 74(5):357–366
- Habib G, Kiryati N, Sklair-Levy M, Shalmon A, Neiman O, Weidenfeld R, Yagil Y, Konen E, Mayer A (2020) Automatic breast lesion classification by joint neural analysis of mammography and ultrasound. In: Multimodal learning for clinical decision support and clinical image-based procedures: 10th international workshop, ML-CDS 2020, and 9th international workshop, CLIP 2020, held in conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings, vol 12445, pp 125–135. Springer
- Shaikh TA, Rashid A, Sufyan Beg MM (2020) Transfer learning privileged information fuels cad diagnosis of breast cancer. *Mach Vis Appl* 31(1):1–23
- Mateos MJ, Gastelum A, Márquez J, Brandan ME (2016) Texture analysis of contrast-enhanced digital mammography (CEDM) images. In: Lecture Notes in Computer Science (including sub-series lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 9699. Springer, pp 585–592
- Shaked P, Nahum K, Gali Z-M, Miri S-L, Eli K, Arnaldo M (2019) Classification of contrast-enhanced spectral mammography (CESM) images. *Int J Comput Assist Radiol Surg* 14(2):249–257
- Laura L, Menell Jennifer H (2002) Breast imaging reporting and data system (BI-RADS). *Radiol Clin North Am* 40(3):409–430
- Fanizzi A, Losurdo L, Basile TMA, Bellotti R, Bottigli U, Delogu P, Diacono D, Didonna V, Fausto A, Lombardi A, Lorusso V, Massafra R, Tangaro S, Forgia DL (2019) Fully automated support system for diagnosis of breast cancer in contrast-enhanced spectral mammography images. *J Clin Med* 8(6):891
- Liliana L, Annarita F, Basile Teresa Maria A, Roberto B, La Forgia D (2019) Radiomics analysis on contrast-enhanced spectral mammography images for breast cancer diagnosis: a pilot study. *Entropy* 21(11):1110
- Daniele LF, Annarita F, Francesco C, Roberto B, Vittorio D, Vito L, Marco M, Raffaella M, Pasquale T, Sabina T, Michele T, Maria P, Alfredo Z (2020) Radiomic analysis in contrast-enhanced spectral mammography for predicting breast cancer histological outcome. *Diagnostics* 10:09

26. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
27. Kaiming H, Xiangyu Z, Shaoqing R, Jian S (2016). Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
28. Jie H, Li S, Gang S (2018). Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
29. Saining X, Ross G, Piotr D, Zhuowen T, Kaiming H (2017). Aggregated residual transformations for deep neural networks. In: Proceedings - 30th IEEE conference on computer vision and pattern recognition, CVPR 2017, vol 2017, pp 5987–5995
30. Zagoruyko S, Komodakis N (2016). Wide residual networks. In: British machine vision conference 2016, BMVC 2016, pp 87.1–87.12
31. Gao S, Cheng M-M, Zhao K, Zhang X-Y, Yang M-H, Torr PHS (2019). Res2net: a new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, pp :1–1, 08
32. Peng Qi, Wei Zhou, Jizhong Han (2017) A method for stochastic L-BFGS optimization, vol 434, pp 156–160
33. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017). Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp 618–626

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