European Radiology (2024) 34:917–927
https://doi.org/10.1007/s00330-023-10170-9

European Radiology
EUROPEAN SOCIETY OF RADIOLOGY

BREAST

Breast cancer diagnosis from contrast-enhanced mammography using multi-feature fusion neural network

Nini Qian[1] · Wei Jiang[1,2] · Yu Guo[1] · Jian Zhu[3] · Jianfeng Qiu[4] · Hui Yu[1] · Xian Huang[1]

Received: 23 December 2022 / Revised: 25 May 2023 / Accepted: 8 July 2023 / Published online: 23 August 2023

Abstract

Objectives To develop an end-to-end deep neural network for the classification of contrast-enhanced mammography (CEM) images to facilitate breast cancer diagnosis in the clinic.

Methods In this retrospective mono-centric study, patients who underwent CEM examinations from January 2019 to August 2021 were enrolled. A multi-feature fusion network combining low-energy (LE) and dual-energy subtracted (DES) images and dual view, as well as bilateral information, was trained and tested using a large CEM dataset with a diversity of breast tumors for breast lesion classification. Its generalization performance was further evaluated on two external datasets. Results were reported using AUC, accuracy, sensitivity, and specificity.

Results A total of 2496 patients (mean age, 53 years ± 12 (standard deviation)) were included and divided into a training set (1718), a validation set (255), and a testing set (523). The proposed CEM-based multi-feature fusion network achieved the best diagnosis performance with an AUC of 0.96 (95% confidence interval (CI): 0.95, 0.97), compared with the no-fusion model, the left-right fusion model, and the multi-feature fusion network with only LE image inputs. Our models reached an AUC of 0.90 (95% CI: 0.85, 0.94) on a full-field digital mammograph (FFDM) external dataset (86 patients), and an AUC of 0.92 (95% CI: 0.89, 0.95) on a CEM external dataset (193 patients).

Conclusion The developed multi-feature fusion neural network achieved high performance in CEM image classification and was able to facilitate CEM-based breast cancer diagnosis.

Clinical relevance statement Compared with low-energy images, CEM images have greater sensitivity and similar specificity in malignant breast lesion detection. The multi-feature fusion neural network is a promising computer-aided diagnostic tool for the clinical diagnosis of breast cancer.

Key Points

- *Deep convolutional neural networks have the potential to facilitate contrast-enhanced mammography-based breast cancer diagnosis.*
- *The multi-feature fusion neural network reaches high accuracies in the classification of contrast-enhanced mammography images.*
- *The developed model is a promising diagnostic tool to facilitate clinical breast cancer diagnosis.*

Keywords Breast neoplasms · Mammography · Deep learning · Contrast media

Nini Qian and Wei Jiang contributed equally to this work.

✉ Yu Guo
guoyu@tju.edu.cn

[1] Department of Biomedical Engineering, School of Precision Instrument and Opto-electronics Engineering, Tianjin University, Tianjin 300072, China

[2] Department of Radiotherapy, Yantai Yuhuangding Hospital, No. 20 Yuhuangding East Road, Yantai 264000, Shandong, China

[3] Department of Radiation Oncology Physics and Technology, Shandong Cancer Hospital, Jiyan Road, Jinan 250117, Shandong, China

[4] Medical Engineering and Technology Research Center, School of Radiology, Shandong First Medical University & Shandong Academy of Medical Sciences, Taian 271000, Shandong, China

Abbreviations

CAM	Class activation map
CC	Craniocaudal
CEM	Contrast-enhanced mammography
CI	Confidence interval
CNN	Convolutional neural network
DES	Dual-energy subtracted
FFDM	Full-field digital mammography
LE	Low-energy
MLO	Mediolateral oblique
NPV	Negative predictive value
PPV	Positive predictive value
VGG	Visual geometry group

Introduction

Nowadays, female breast cancer has become the most diagnosed cancer worldwide in 2020, with approximately 2.3 million new cases reported [1]. In China, breast cancer has been the leading cancer type in women, accounting for 18% of all global breast cancer cases [2].

Contrast-enhanced mammography (CEM) is a promising breast imaging technique that utilizes iodine contrast agents and dual-energy mammography to show angiogenesis and visualize abnormal breast lesions. Because of the improved ability of CEM to detect tumors shielded by dense breast glands compared with mammography, CEM has great feasibility and potential in initial breast cancer screening [3–10]. Recent studies have shown that CEM had a higher positive predictive value (PPV) and specificity compared with MRI for cancer detection and outperformed full-field digital mammography (FFDM) in terms of diagnostic accuracy [11–13]. Although other researchers have demonstrated that the sensitivity of detecting malignant breast lesions had no significant difference between CEM and MRI [14–16], CEM was popular with patients due to its faster procedure time, more comfortable patient positioning, lower rate of anxiety, and less cost [17]. Altogether, CEM is playing significant roles not only in breast cancer screenings but also in the assessment of symptomatic patients and the preoperative staging of breast cancer [18, 19].

To assist radiologists in improving CEM-based breast cancer diagnosis, many computer-aided methods have been proposed in the literature. Some researchers carried out radiomics analyses on lesion regions manually delineated in CEM images [20–22]. Gao et al [23] proposed a shallow-deep convolutional neural network for predicting whether breast lesions annotated by bounding boxes in CEM images were benign or malignant. However, methods in the literature required lesion delineations and were conducted mostly on small CEM datasets, lacking diversity and complexity in lesions.

In clinical practice, CEM-based breast cancer detection is performed by combining dual-view, low-energy (LE), and dual-energy subtracted (DES) images of bilateral breasts, which is supposed to facilitate the diagnosis. Integrating information manually from multiple images requires effort and extensive breast imaging experience. In this paper, an end-to-end deep multi-feature fusion neural network, which combined the information of bilateral, dual-view, and dual-energy mammography images, was proposed for the automatic diagnosis of breast cancer based on CEM images.

Materials and methods

Patient selection

This retrospective study was approved by the Institutional Ethics Committee of the Yantai Yuhuangding Hospital (Approval Number: 2023-073), and the informed consent requirement was waived. All patients who underwent CEM examinations at the Yantai Yuhuangding Hospital from January 2019 to August 2021 were initially included. Indications for CEM were the assessment of suspicious lesions detected at screening, the evaluation of symptomatic women, and post-therapy monitoring. In our study, we excluded the following cases: (1) with suspicious findings but without pathological reports; (2) with a history of breast surgery, radiotherapy, or chemotherapy. The detailed patient exclusion criteria are shown in Fig. 1.

Imaging acquisition

All diagnostic CEM examinations were performed on the GE Senographe Essential Mammography System. Patients were first intravenously administered the iodine contrast

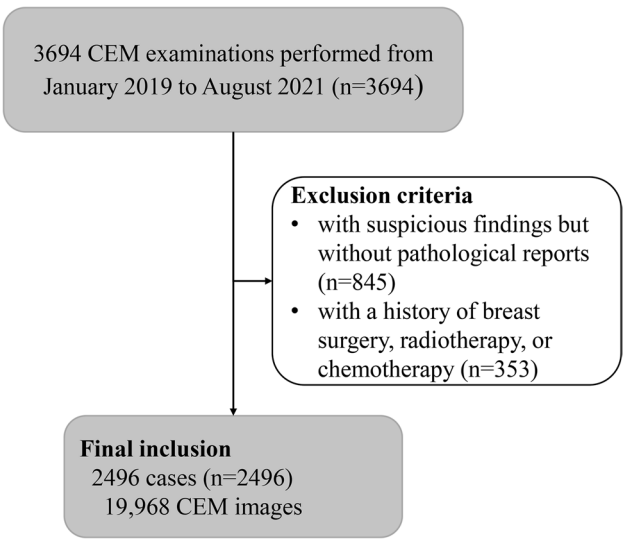

Fig. 1 Flow diagram of the study exclusion criteria

agent injection (a dose of 1.5 mL/kg body weight, an injection flow rate of 3.0 mL/s) before image acquisition. Two minutes later, the low-energy images of both breasts in craniocaudal (CC) and mediolateral oblique (MLO) views were obtained at a low-energy exposure of 26–32 kilovoltage (kVp). Almost simultaneously, the high-energy images were acquired at a high-energy exposure of 45–49 kVp. Then, the dual-energy subtracted (DES) images were generated from the low-energy and high-energy images by using the postprocessing algorithm [24] to highlight contrast enhancement areas. Finally, each CEM examination provided eight images (LE and DES images of two breasts in CC and MLO views) with a size of 3062 × 2394. After the examination, each patient was given a Breast Imaging Reporting and Data System (BI-RADS) assessment category to describe her CEM examination result. Suspicious findings were further proven by biopsy.

Data division

The CEM dataset was randomly divided into a training, validation, and testing set with a ratio of 7:1:2. Neural network models were trained on the training set. When the prediction errors on the validation set began to increase, network training was stopped, and then, the fitted models were evaluated in the testing set.

Data preprocessing and augmentation

The pixel values of each image were adjusted and limited to the range of 1500 to 3300 to enhance the contrast of CEM images and highlight lesion regions. Also, the useless background was cropped out, and min-max normalization was carried out. Finally, preprocessed CEM images were resized to 512 × 256 with bilinear interpolation. Online data augmentation was also utilized. CEM images were randomly

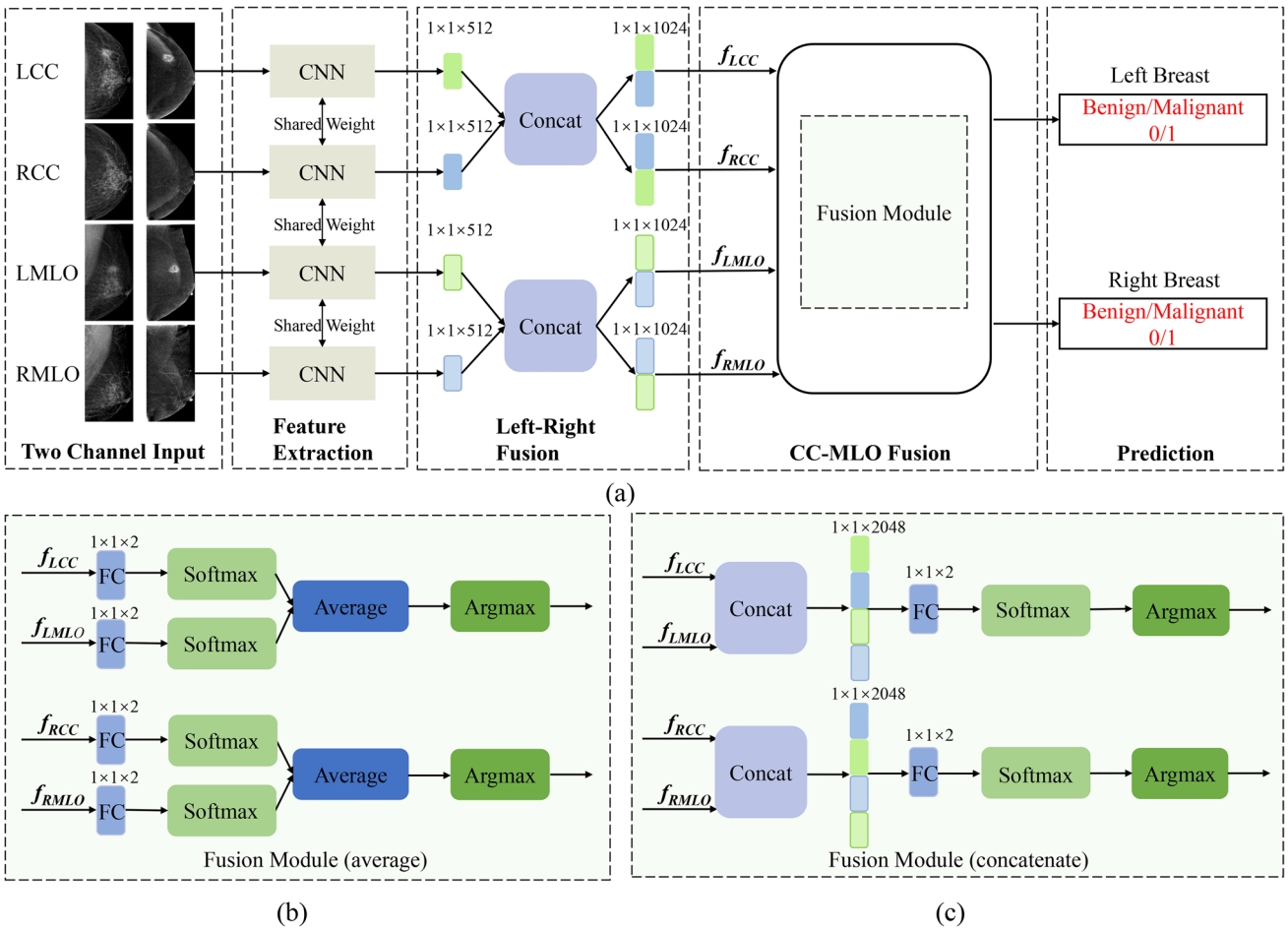

Fig. 2 Multi-feature fusion network architecture. **a** General framework of the multi-feature fusion network. **b** Average fusion module for dual-view fusion. **c** Concatenate fusion module for dual-view fusion. LCC, dual-energy images of the left breast in craniocaudal (CC) view; LMLO, dual-energy images of the left breast in mediolateral oblique (MLO) view; RCC, dual-energy images of the right breast in CC view; RMLO, dual-energy images of the right breast in MLO view; f_{LCC}, left-right fusion features of the left breast in CC view; f_{LMLO}, left-right fusion features of the left breast in MLO view; f_{RCC}, left-right fusion features of the right breast in CC view; f_{RMLO}, left-right fusion features of the right breast in MLO view; CNN, convolutional neural network; FC, fully connected layers

flipped horizontally and vertically with a probability of 0.5 and rotated randomly in the range of [−10°, 10°].

Deep learning model

In this paper, we developed a multi-feature fusion neural network for breast cancer detection by combining information from CEM images of dual views of bilateral breasts. The general framework of the proposed network is shown in Fig. 2a. Each pair of LE and DES images was concatenated together and regarded as the two-channel input of a convolutional neural network (CNN), which extracted the fused features of the dual-energy images. Accordingly, a total of four CNNs, sharing the same weights, were employed to extract features from the images in CC and MLO views of both breasts of the same patient. To combine the contralateral breast information for the diagnosis of a certain breast, the extracted features of the bilateral breasts in the same view were firstly concatenated to create a single feature vector. The fused feature vectors of CC and MLO views were further combined in the CC-MLO fusion module.

Here, two fusion strategies were proposed for CC and MLO view fusion. One was a decision fusion approach, as shown in Fig. 2b. The features of each view were inputted in a fully connected layer with softmax output to predict the benign and malignant probabilities of a breast, and then, the prediction results based on different views were fused with an average operation to output the final benign and malignant probability of the breast. Eventually, the diagnosis output was the class with the largest predicted probability determined by an argmax operation. The other fusion strategy was a feature concatenation fusion approach (Fig. 2c), in which the feature vectors of different views were concatenated as a feature vector to be fed into a binary image classifier, which was composed of a fully connected layer, a softmax, and an argmax operation.

About the CNN architecture, a visual geometry group with a depth of 16 layers (VGG16) [25] was used in this study. The detailed architecture of VGG16 is illustrated in Fig. 3.

The CEM training set was represented as $\{x_n, y_n, n = 1, 2, 3, ..., N\}$,where N was the patient number in the training set, x_n was the CEM images of the nth patient, and y_n was the ground truth label of that patient, indicating whether each breast contained malignant tumors. Then, the cross-entropy loss function denoted as follows was employed to train the proposed network.

$$\text{Loss} = \frac{1}{N}\sum_{n=1}^{N} H(\hat{y}_n, y_n), \tag{1}$$

$$H(\hat{y}_n, y_n) = -\left(y_n \log(\hat{y}_n) + (1 - y_n)\log(1 - \hat{y}_n)\right), \tag{2}$$

where $\hat{y}_n$ denoted the probability that the sample was predicted to be positive.

Model training

Concerning deep neural network-based classification, network training strategies had an important influence on the final classification performance. In this study, the CNN modules were initially trained with all the LE and DES image pairs in the training set to be capable of extracting the relevant image features. Afterward, the trained CNN model was fine-tuned in the final fusion framework.

The SGD optimizer with a weight decay factor of 5×10^{-4} and momentum of 0.9 was used while training. All the experiments were implemented on an Intel(R) Core(TM) i7-9700 CPU @ 3.00 GHz and an NVIDIA GeForce RTX 3090 24 GB GPU. The software environments were Ubuntu 18.04 and Pytorch 3.9. The batch size was set to 8. The initial learning rate was 0.001. Once the training epoch reached the corresponding settings (10th, 30th, and 40th epoch), the learning rate decayed to half of the previous rate. The code of the

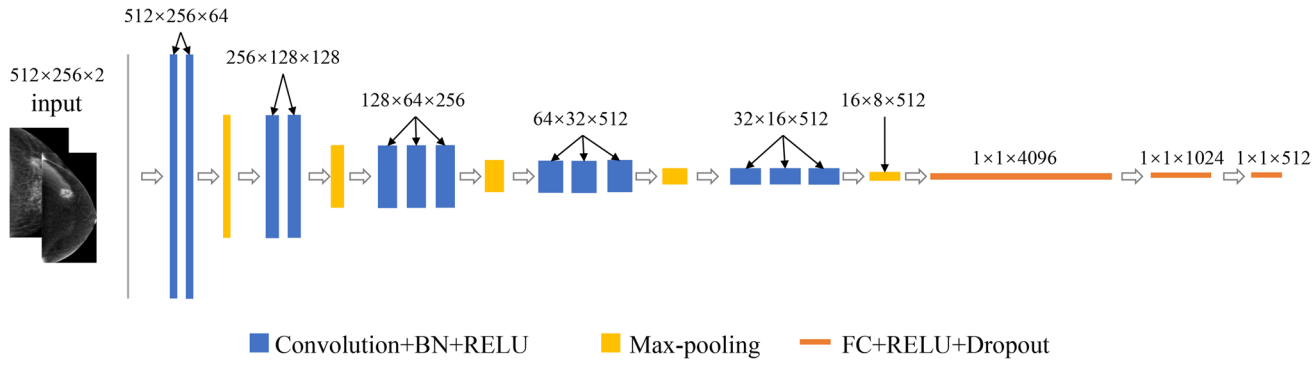

Fig. 3 Architecture of convolutional neural network visual geometry group 16 (VGG16), which was composed of convolution layers, batch normalization layers, rectified linear units, max-pooling layers, and fully connected layers. Feature size = height × width × channel. BN, batch normalization; RELU, rectified linear unit; FC, a fully connected layer

multi-feature fusion model is publicly available (https://github.com/anne226/MF_MODEL.git).

Statistical analysis

To assess the classification performance of our models, several metrics, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under the receiver operating characteristic (ROC) curve (AUC), were utilized.

Delong's test [26] was used to compare the significance level of ROC curves among different models. p values < 0.05 were considered a statistically significant difference. Statistical analyses were performed with the MedCalc application (version 19.6.4-64bit, https://www.medcalc.org/) and with R software (version 4.2.1-64bit, https://www.r-project.org/). The R packages used in this study included pROC, caret, openxlsx, epiR, and ggplot2.

Results

Patient characteristics

A total of 2496 female patients (mean age, 53 years ± 12 standard deviation (SD); range, 17–85 years) were finally enrolled in this study. Detailed patient characteristics are shown in Table 1. These eligible patients were randomly allocated into a training set (1718 patients, mean age, 53 years ± 12 (SD)), a validation set (255 patients, mean age, 53 years ± 11 (SD)), and a testing set (523 patients, mean age, 52 years ± 12 (SD)). More details about the dataset splitting are shown in Table 2.

Table 1 Patient characteristics: age, BI-RADS categories, and biopsy results

Age	Years
Range	17–85
Mean ± standard deviation	53 ± 12
BI-RADS	Percentage ($n = 2496 \times 2$)
0	0.8
1	14.2
2	3
3	30.5
3–4A	2.3
4A	11.4
4B	8.4
4C	11.8
5	12
6	5.6
Total	100
Breast biopsy results (left/right breast)	Percentage ($n = 2496 \times 2$)
Begin	64.5
Malignant	35.5
DCIS	1.9
IDC	18.5
DCIS + IDC	10.5
LCIS	0.1
ILC	0.8
LCIS + ILC	0.2
Papillary carcinoma	1.9
Mucinous carcinoma	1.1
Malignant phyllodes tumor	0.2
Paget's disease	0.2
Squamous cell carcinoma	0.1
Total	100

BI-RADS, Breast Imaging Reporting and Data System; *DCIS*, ductal carcinoma in situ; *IDC*, invasive ductal carcinoma; *LCIS*, lobular carcinoma in situ; *ILC*, invasive lobular carcinoma. At the time of statistics, special types of breast cancer were partly accompanied by common breast cancer types (ductal carcinoma in situ/invasive ductal carcinoma)

Performance on the internal testing set

The proposed multi-feature fusion network using the feature concatenation operation for CC and MLO view fusion achieved a classification accuracy of 0.90 and an AUC of 0.96 in the testing set. Forty-seven benign tumors (4.5%) were misclassified as malignant, while 59 malignant tumors (5.6%) were incorrectly classified as benign. Among the incorrectly classified malignant cases, 31 (31/59) were extremely dense breasts, and 9 breasts (9/51) were diagnosed as special types of breast cancer, such as mucinous carcinoma, malignant phyllodes tumor, and solid papillary carcinoma. The remaining malignant lesions that were incorrectly classified were either too small or not visibly enhanced in CEM images.

As for different BI-RADS assessment categories, the multi-feature fusion network got an overall classification accuracy of 98%, a sensitivity of 96%, and a specificity of 99% for categories 1, 2, 3, and 5. However, its classification accuracy for category 4 (including 3–4A, 4A, 4B, and 4C) was 80%, and the sensitivity was 81%.

The performance of the multi-feature fusion network was further compared to that of no-fusion models, left-right fusion models, and models with only LE image inputs, as shown in Tables 3 and 4 and Fig. 4. No-fusion models only used a CNN to classify input images, while left-right fusion models classified input images based on the feature vector formed by the concatenation of the CNN features of a breast and its contralateral breast in the same view. CNN modules used to extract image features in all these models had the same architecture.

Compared with the no-fusion model, the CEM-based left-right fusion model improved prediction AUC from 0.93 to 0.95. By combining images of not only bilateral breasts but also dual views, the proposed multi-feature fusion methods

Table 2 Standardized training/validation/testing split in the internal contrast-enhanced mammography dataset

	Bilateral benign	Unilateral malignant	Bilateral malignant	Overall
Training	542 (31.5%)	1149 (66.9%)	27 (1.6%)	1718 (100%)
Validation	65 (25.5%)	185 (72.5%)	5 (2%)	255 (100%)
Testing	158 (30.2%)	358 (68.4%)	7 (1.4%)	523 (100%)
Overall	765 (30.6%)	1692 (67.8%)	39 (1.6%)	2496 (100%)

furtherly improved the prediction performance. Besides, neither of the two CC-MLO fusion strategies outperformed the other when all evaluation metrics were considered.

When only LE images were used for breast cancer diagnosis, the multi-feature fusion model using the concatenate operation had a lower classification accuracy of 88% and an AUC of 0.94 in the testing set compared to the models with both LE and DES inputs.

Delong's test was applied to evaluate the discrimination in different models. The analysis results (Table 5) showed that there was a significant difference between the no-fusion model and the left-right fusion model ($p < 0.001$) and between the models trained on CEM images (DES + LE) and only on LE images ($p = 0.003$ for multi-feature fusion by the average operation and $p = 0.006$ for multi-feature fusion by the concatenate operation). However, there was no significant difference between the two CC and MLO fusion methods ($p = 0.91$ for LE input and $p = 0.61$ for LE and DES inputs).

Class activation maps

The generalized gradient-based class activation mapping [27] technique is a visualized tool for illustrating the contributions of every pixel to the final classification output of a CNN model. Here, it was used to generate the class activation maps (CAMs) of each breast to visualize the regions that different networks were interested in.

Figure 5a illustrates the CAMs of a breast with a malignant tumor generated by different models, including the no-fusion model, left-right fusion model, and multi-feature fusion

Table 3 Performance evaluation of different models with low-energy image inputs

	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)	PPV (95% CI)	NPV (95% CI)
LE (no-fusion)	0.90 (0.88, 0.91)	0.85 (0.83, 0.87)	0.74 (0.70, 0.77)	0.91 (0.90, 0.93)	0.82 (0.79, 0.85)	0.86 (0.84, 0.88)
LE (left-right fusion)	0.93 (0.92, 0.94)	0.86 (0.85, 0.88)	0.75 (0.72, 0.78)	0.92 (0.91, 0.94)	0.84 (0.81, 0.87)	0.87 (0.85, 0.89)
LE (multi-feature fusion (average))	0.94 (0.92, 0.95)	0.86 (0.84, 0.88)	0.74 (0.69, 0.79)	0.93 (0.91, 0.95)	0.86 (0.81, 0.89)	0.87 (0.84, 0.89)
LE (multi-feature fusion (concatenate))	0.94 (0.93, 0.95)	0.88 (0.86, 0.90)	0.79 (0.74, 0.83)	0.93 (0.91, 0.95)	0.86 (0.82, 0.89)	0.89 (0.86, 0.91)

CI, confidence interval; *AUC*, area under the receiver operating characteristic curve; *ACC*, accuracy; *SEN*, sensitivity; *SPE*, specificity; *PPV*, positive predictive value; *NPV*, negative predictive value; *LE*, low-energy; *CEM*, contrast-enhanced mammography

Table 4 Performance evaluation of different models with contrast-enhanced mammography image inputs

	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)	PPV (95% CI)	NPV (95% CI)
CEM (no-fusion)	0.93 (0.92, 0.94)	0.87 (0.85, 0.88)	0.77 (0.74, 0.80)	0.92 (0.91, 0.94)	0.85 (0.82, 0.87)	0.88 (0.86, 0.89)
CEM (left-right fusion)	0.95 (0.94, 0.96)	0.88 (0.87, 0.89)	0.80 (0.76, 0.82)	0.93 (0.91, 0.94)	0.86 (0.83, 0.88)	0.89 (0.87, 0.91)
CEM (multi-feature fusion (average))	0.96 (0.95, 0.97)	0.89 (0.87, 0.91)	0.85 (0.81, 0.88)	0.92 (0.89, 0.94)	0.85 (0.81, 0.88)	0.92 (0.89, 0.94)
CEM (multi-feature fusion (concatenate))	0.96 (0.95, 0.97)	0.90 (0.88, 0.92)	0.84 (0.80, 0.88)	0.93 (0.91, 0.95)	0.87 (0.83, 0.90)	0.91 (0.89, 0.93)

CI, confidence interval; *AUC*, area under the receiver operating characteristic curve; *ACC*, accuracy; *SEN*, sensitivity; *SPE*, specificity; *PPV*, positive predictive value; *NPV*, negative predictive value; *LE*, low-energy; *CEM*, contrast-enhanced mammography

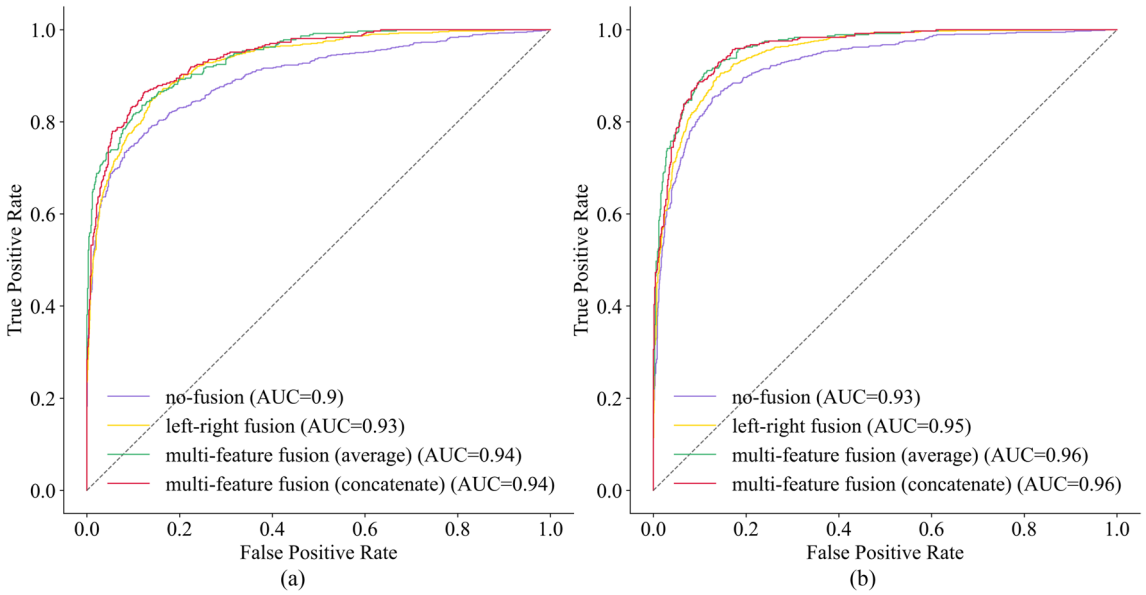

Fig. 4 Receiver operating characteristic curve. **a** Models with low-energy image inputs. **b** Models with contrast-enhanced mammography image inputs

model using the concatenate operation, which were all with CEM (DES + LE) image inputs. Compared with the no-fusion model and left-right fusion model, the multi-feature fusion model was able to concentrate more clearly on the lesions.

We also analyzed the CAMs generated by models with different inputs. As shown in Fig. 5b, the model with CEM image inputs focused on the benign tumor and achieved the correct diagnosis, while the model with only LE inputs did not pay attention to the whole tumor and made an incorrect prediction.

Performance on external test datasets

To evaluate the generalization performance of the proposed multi-feature fusion network on external datasets, a public FFDM database, INbreast [28], and a public CEM database [29] were used.

The INbreast dataset was acquired from the breast center in Centro Hospitalar de São João (CHSJ) at Porto, which contained 115 cases (410 images) in total. The masks of ROIs were annotated and validated by two specialists. In our study, only cases (a total of 86 cases) with bilateral FFDM images were used, which meant that each case had 4 FFDM images. As biopsy-proven annotations were not available, these cases were labeled as benign or malignant according to BI-RADS results. BI-RADS categories 1, 2, and 3 were regarded as benign, while the assessment categories 4, 5, and 6 were regarded as breast cancer. Finally, 44 (51.2%) cases were assigned as bilateral benign tumors, 41 (47.7%) cases were assigned as unilateral breast malignancies, and only 1 (1.1%) case was considered bilateral malignancy.

As FFDM is a mono-energy imaging technique, the trained multi-feature fusion model on the internal dataset using the concatenating operation with LE image inputs was used for the classification of FFDM images in INbreast. The same data preprocessing technique as for CEM images was utilized for FFDM image preprocessing. Finally, we got an

Table 5 Significance levels of different models

Model 1	Model 2	*p* value
LE (no-fusion)	LE (left-right fusion)	< .001
LE (multi-feature fusion (average))	LE (multi-feature fusion (concatenate))	.91
CEM (no-fusion)	CEM (left-right fusion)	< .001
CEM (multi-feature fusion (average))	CEM (multi-feature fusion (concatenate))	.61
LE (no-fusion)	CEM (no-fusion)	< .001
LE (multi-feature fusion (average))	CEM (multi-feature fusion (average))	.003
LE (left-right fusion)	CEM (left-right fusion)	< .001
LE (multi-feature fusion (concatenate))	CEM (multi-feature fusion (concatenate))	.006

LE, low-energy; *CEM*, contrast-enhanced mammography. LE represented models with low-energy image inputs. CEM represented models with contrast-enhanced mammography image inputs. *p* values were calculated by using Delong's test

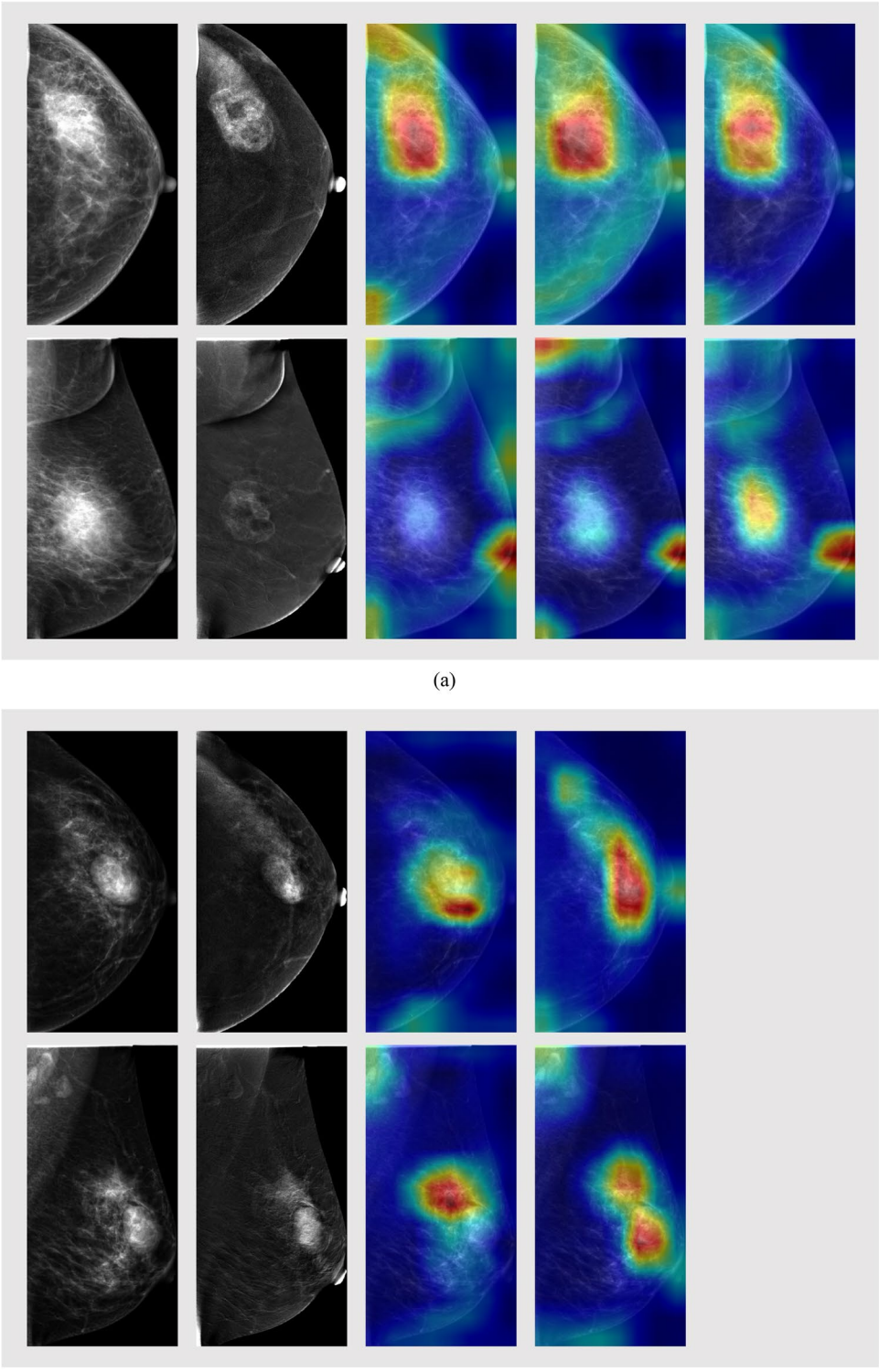

Fig. 5 Visualized class activation maps (CAMs) results of two cases from the internal testing set. **a** Unilateral breast in craniocaudal view and mediolateral oblique view of a 57-year-old female biopsy-proven to be invasive ductal carcinoma. Left to right: low-energy images, dual-energy subtracted images, and CAMs based on the no-fusion model, the left-right fusion model, and the multi-feature fusion model by the concatenate operation. **b** Unilateral breast in craniocaudal view and mediolateral oblique view of a 31-year-old female diagnosed with benign phyllodes tumors. Left to right: low-energy images, dual-energy subtracted images, and CAMs based on the multi-feature fusion model by the concatenate operation with low-energy image or contrast-enhanced mammography image inputs

AUC of 0.90 (95% CI: 0.85, 0.94), an accuracy of 0.84 (95% CI: 0.78, 0.89), a sensitivity of 0.77 (95% CI: 0.61, 0.88), and a specificity of 0.87 (95% CI: 0.80, 0.92). As shown in Fig. 6a, the proposed multi-feature fusion network was able to precisely focus on the lesion areas in FFDM images.

A CEM dataset gathered from the Radiology Department of the National Cancer Institute at Cairo University was released in [29]. This dataset contained images of 326 female patients, 193 of whom had images of bilateral breasts in both CC and MLO views. Images from these 193 patients were used in this

study as the external CEM dataset. Pathological results and follow-ups were provided in the annotations. Sixty-two (32.1%) cases were considered bilateral benign tumors, 127 (65.8%) cases were assigned as unilateral breast malignancies, and only 4 (2.1%) cases were considered bilateral malignancies.

The multi-feature fusion network with CEM inputs trained on our internal training set was used to classify CEM images in this external dataset. The classification results showed an AUC of 0.92 (95% CI: 0.89, 0.95), an accuracy of 0.85 (95% CI: 0.82, 0.89), a sensitivity of 0.86 (95% CI: 0.79, 0.91), and a specificity of 0.85 (95% CI: 0.80, 0.89). Figure 6b shows the final CAMs of one case, which indicate the location of a breast lesion.

Discussion

In this study, a convolutional neural network-based multi-feature fusion method was proposed for breast cancer diagnosis based on a large amount of contrast-enhanced mammography (CEM) data. Compared with the no-fusion model and left-right fusion model, the proposed multi-feature

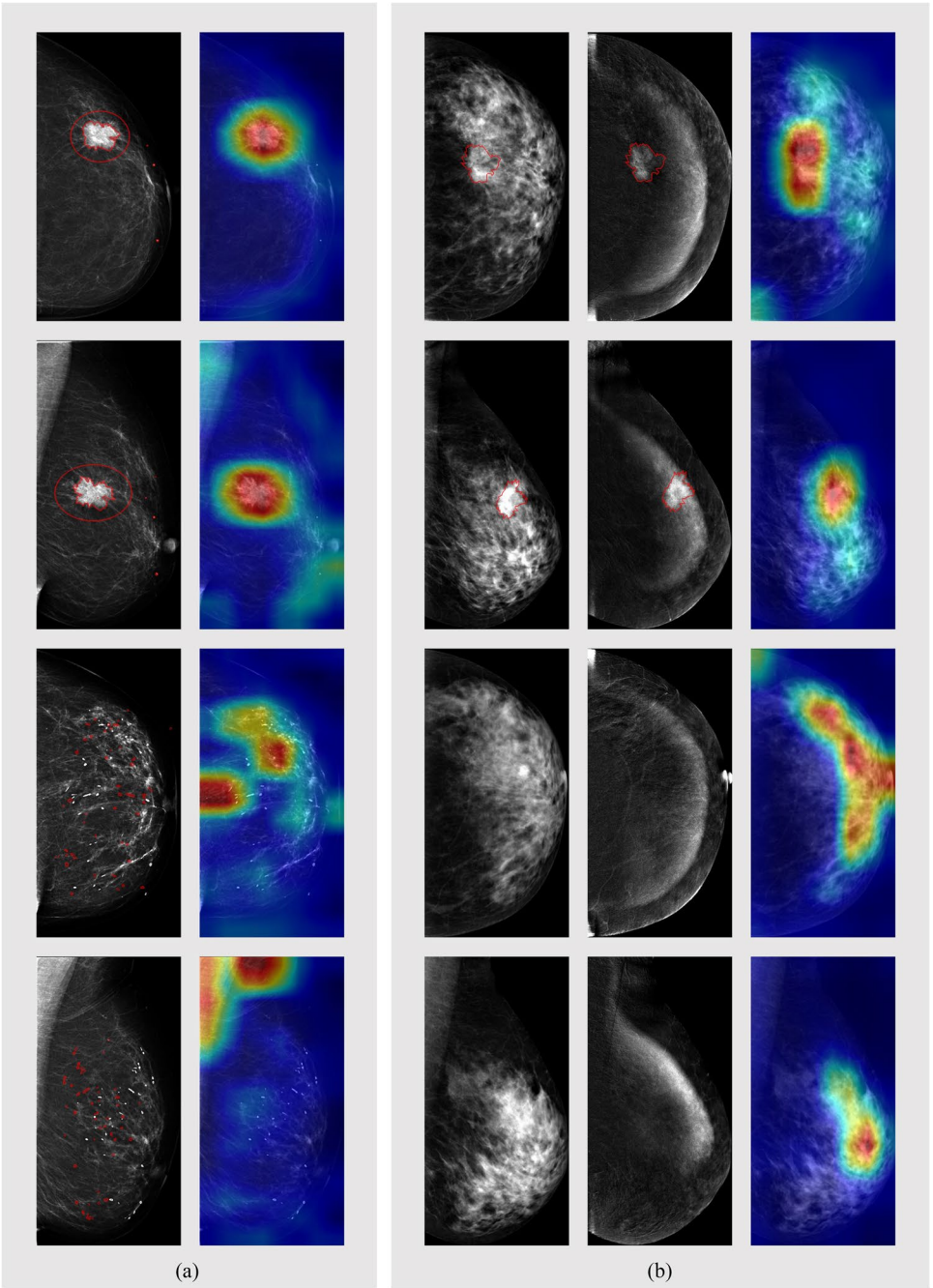

Fig. 6 Visualized class activation maps (CAMs) results of two cases from the external test dataset. **a** FFDM images and CAMs of a patient from INbreast dataset (left breast: BI-RADS category 5; right breast: category 3). **b** Low-energy images, dual-energy subtracted images, and the CAMs of a patient from CDD-CEM dataset (left breast: malignant; right breast: normal). The lesions were highlighted with red points and circles. Top to bottom: craniocaudal view and mediolateral oblique view of the left breast and of the right breast

fusion model achieved better diagnosis performance with an AUC of 0.96 (95% confidence interval (CI): 0.95, 0.97) and an accuracy of 0.90 (95% CI: 0.88, 0.92). The improved performances of the left-right fusion models demonstrated that the feature fusion of bilateral breasts enabled the networks to learn the most discriminative classification features of a breast by comparing its image features to those of its contralateral breast. Dual-view fusion further increased the accuracy of deep neural networks in CEM-based breast cancer diagnosis. Besides, this study illustrated that CEM-based breast cancer diagnosis models outperformed low-energy (LE) mammogram-based models on breast cancer diagnosis, especially in sensitivity. Furthermore, the generalization performance of the developed multi-feature fusion model was evaluated on two external datasets: a public full-field digital mammography (FFDM) dataset, INbreast, and a public CEM dataset. The test results demonstrated the good generalization performance of our models in external datasets collected from different institutions and countries.

Many methods have been proposed for breast lesion classification in CEM images in the literature. Fanizzi et al [20] have proposed an automatic computer-automated diagnosis system for the classification of breast lesions delineated manually in CEM images. In this method, image features were extracted from lesion regions and the random forest classifier was trained to classify lesions as benign or malignant tumors. Testing on 48 lesions, this method resulted in an AUC of 0.931, an accuracy of 87.5%, a sensitivity of 87.5%, and a specificity of 91.7%. Losurdo et al [21] developed a lesion classification model based on 55 breast lesions of 51 patients. Using significant CEM features extracted from lesion regions reached the best classification accuracy of 80%. Patel et al [22] studied 50 lesions in CEM images, from which textural and morphologic features were extracted. Using the support vector machine classifier, they finally achieved a 0.95 AUC, 90% accuracy, 88% sensitivity, and 92% specificity. Gao et al [23] developed a shallow convolutional neural network based on 49 CEM cases to generate virtual recombined images, and a CEM classification model for breast cancer diagnosis, in which a convolutional neural network (CNN) was used to extract image features and gradient boosting trees were used for the final classification. This study attained a 0.92 AUC, 90% accuracy, 83% sensitivity, and 94% specificity.

However, our study still has several limitations. The proposed models were trained on the CEM images collected from a single institution. A highly standardized, diverse, large, and multicentric CEM database is required for further research. Besides, our model only got an accuracy of 80% for Breast Imaging Reporting and Data System (BI-RADS) category 4 lesions. In further work, focusing on the malignancy prediction of microcalcifications will enhance the diagnosis accuracy of models for these lesions. Furthermore, if pixel-level labels can also be provided by specialists in addition to breast-level labels, the prediction accuracy will be further improved. Analyses of the image features extracted by deep CNN will also help to understand how deep neural network-based models assist in breast cancer diagnosis.

In conclusion, the developed multi-feature fusion neural network trained on a large and diverse dataset achieved high accuracy in breast tumor detection. It has the potential to be used in the clinic for facilitating CEM-based breast tumor diagnosis.

Funding This study has received funding by Major Science and Technology Projects in Tianjin (18ZXZNSY00240), Science and Technology Innovation Projects for Medical System Staff in Shandong Province (SDYWZGKCJH2022042).

Declarations

Guarantor The scientific guarantor of this publication is Yu GUO, Ph.D., Department of Biomedical Engineering, School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin, China.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Ethics Committee of the Yantai Yuhuangding Hospital.

Ethical approval This study was approved by the Institutional Ethics Committee of the Yantai Yuhuangding Hospital (Approval Number: 2023-073).

Study subjects or cohorts overlap Our study cohorts have not been previously reported.

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/caac.21660
2. Cao W, Chen HD, Yu YW, Li N, Chen WQ (2021) Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. Chin Med J 134:783–791. https://doi.org/10.1097/cm9.0000000000001474
3. Jong RA, Yaffe MJ, Skarpathiotakis M et al (2003) Contrast-enhanced digital mammography: initial clinical experience. Radiology 228:842–850. https://doi.org/10.1148/radiol.2283020961

4. Lewin JM, Isaacs PK, Vance V, Larke FJ (2003) Dual-energy contrast-enhanced digital subtraction mammography: feasibility. Radiology 229:261–268. https://doi.org/10.1148/radiol.2291021276
5. Fallenberg EM, Schmitzberger FF, Amer H et al (2017) Contrast-enhanced spectral mammography vs. mammography and MRI - clinical performance in a multi-reader evaluation. Eur Radiol 27:2752–2764. https://doi.org/10.1007/s00330-016-4650-6
6. Cheung YC, Lin YC, Wan YL et al (2014) Diagnostic performance of dual-energy contrast-enhanced subtracted mammography in dense breasts compared to mammography alone: interobserver blind-reading analysis. Eur Radiol 24:2394–2403. https://doi.org/10.1007/s00330-014-3271-1
7. Sorin V, Yagil Y, Yosepovich A et al (2018) Contrast-enhanced spectral mammography in women with intermediate breast cancer risk and dense breasts. AJR Am J Roentgenol 211:W267–W274. https://doi.org/10.2214/ajr.17.19355
8. Lalji U, Lobbes M (2014) Contrast-enhanced dual-energy mammography: a promising new imaging tool in breast cancer detection. Womens Health (Lond) 10:289–298. https://doi.org/10.2217/whe.14.18
9. Lobbes MBI, Smidt ML, Houwers J, Tjan-Heijnen VC, Wildberger JE (2013) Contrast enhanced mammography: techniques, current results, and potential indications. Clin Radiol 68:935–944. https://doi.org/10.1016/j.crad.2013.04.009
10. Lobbes MBI, Lalji U, Houwers J et al (2014) Contrast-enhanced spectral mammography in patients referred from the breast cancer screening programme. Eur Radiol 24:1668–1676. https://doi.org/10.1007/s00330-014-3154-5
11. Lee-Felker SA, Tekchandani L, Thomas M et al (2017) Newly diagnosed breast cancer: comparison of contrast-enhanced spectral mammography and breast MR imaging in the evaluation of extent of disease. Radiology 285:389–400. https://doi.org/10.1148/radiol.2017161592
12. del Mar Travieso-Aja M, Maldonado-Saluzzi D, Naranjo-Santana P et al (2019) Diagnostic performance of contrast-enhanced dual-energy spectral mammography (CESM): a retrospective study involving 644 breast lesions. Radiol Med 124:1006–1017. https://doi.org/10.1007/s11547-019-01056-2
13. Jochelson MS, Dershaw DD, Sung JS et al (2013) Bilateral contrast-enhanced dual-energy digital mammography: feasibility and comparison with conventional digital mammography and MR imaging in women with known breast carcinoma. Radiology 266:743–751. https://doi.org/10.1148/radiol.12121084
14. Sumkin JH, Berg WA, Carter GJ et al (2019) Diagnostic performance of MRI, molecular breast imaging, and contrast-enhanced mammography in women with newly diagnosed breast cancer. Radiology 293:531–540. https://doi.org/10.1148/radiol.2019190887
15. Chou CP, Lewin JM, Chiang CL et al (2015) Clinical evaluation of contrast-enhanced digital mammography and contrast enhanced tomosynthesis-comparison to contrast-enhanced breast MRI. Eur J Radiol 84:2501–2508. https://doi.org/10.1016/j.ejrad.2015.09.019
16. Youn I, Choi S, Choi YJ et al (2019) Contrast enhanced digital mammography versus magnetic resonance imaging for accurate measurement of the size of breast cancer. Br J Radiol 92:20180929. https://doi.org/10.1259/bjr.20180929
17. Hobbs MM, Taylor DB, Buzynski S, Peake RE (2015) Contrast-enhanced spectral mammography (CESM) and contrast enhanced MRI (CEMRI): patient preferences and tolerance. J Med Imaging Radiat Oncol 59:300–305. https://doi.org/10.1111/1754-9485.12296
18. Jochelson MS, Lobbes MBI (2021) Contrast-enhanced mammography: state of the art. Radiology 299:36–48. https://doi.org/10.1148/radiol.2021201948
19. Zhu X, Huang JM, Zhang K et al (2018) Diagnostic value of contrast-enhanced spectral mammography for screening breast cancer: systematic review and meta-analysis. Clin Breast Cancer 18:e985–e995. https://doi.org/10.1016/j.clbc.2018.06.003
20. Fanizzi A, Losurdo L, Basile TMA et al (2019) Fully automated support system for diagnosis of breast cancer in contrast-enhanced spectral mammography images. J Clin Med 8:891. https://doi.org/10.3390/jcm8060891
21. Losurdo L, Fanizzi A, Basile TMA et al (2019) Radiomics analysis on contrast-enhanced spectral mammography images for breast cancer diagnosis: a pilot study. Entropy 21:1110. https://doi.org/10.3390/e21111110
22. Patel BK, Ranjbar S, Wu T et al (2018) Computer-aided diagnosis of contrast-enhanced spectral mammography: a feasibility study. Eur J Radiol 98:207–213. https://doi.org/10.1016/j.ejrad.2017.11.024
23. Gao F, Wu T, Li J et al (2018) SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Comput Med Imaging Graph 70:53–62. https://doi.org/10.1016/j.compmedimag.2018.09.004
24. Jeukens CRLPN (2019) Physics of contrast-enhanced mammography. In: Lobbes M, Jochelson MS, eds Contrast-enhanced mammography. Springer 23–39. https://doi.org/10.1007/978-3-030-11063-5_2
25. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
26. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845. https://doi.org/10.2307/2531595
27. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN (2018) Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE Lake Tahoe, NV. https://doi.org/10.1109/WACV.2018.00097
28. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012) INbreast: toward a full-field digital mammographic database. Acad Radiol 19:236–248. https://doi.org/10.1016/j.acra.2011.09.014
29. Khaled R, Helal M, Alfarghaly O et al (2022) Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. Sci Data 9:122. https://doi.org/10.1038/s41597-022-01238-0