



Recognizing breast tumors based on mammograms combined with pre-trained neural networks

Yujie Bai¹ · Min Li² · Xiaojian Ma³ · Xiaojing Gan³ · Cheng Chen¹ · Chen Chen^{2,5} · Xiaoyi Lv^{1,4,5,6} · Hongtao Li³

Received: 27 August 2021 / Revised: 11 April 2022 / Accepted: 4 February 2023 /

Published online: 11 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Breast cancer is one of the most common cancers in women worldwide, and it seriously threatens people's lives and health. Breast Imaging Reporting and Data System is developed as a standardized system or tool for reporting breast mammograms, where different grades of diagnosis and treatment are critical to the survival rate and survival time of patients. Efficient computer-aided diagnosis of breast tumors based on computer vision models can better assist physicians in selecting effective treatment options, thereby reducing patient mortality. Therefore, early detection and early treatment are of great significance to patients with breast disease. In this study, a new image enhancement framework, called Image Negatives and Contrast Limited Adaptive Histogram Equalization Image Enhancement, was created for the first time based on the comparison of a set of multiple data preprocessing methods for detecting normal, benign, and probably benign breasts. The ResNet-50 pre-trained neural network was used for feature extraction and the classification results were compared on K-nearest neighbor, Random Forest, and Support Vector Machine classifiers. The evaluation indexes adopted in this paper include confusion matrix, precision, sensitivity, F1 Score, etc. These evaluation indexes can be used to evaluate the model in a very comprehensive and accurate way. The experiments show that the KNN classifier has the best classification result, the classification accuracy is 85%, and the AUC is 0.89. It is proved that mammography, as a non-invasive screening tool, has certain practical significance in effectively evaluating tumor grade and its clinical application.

Keywords Breast imaging reporting and data system · Computer-aided diagnosis · Preprocessing · Pre-trained neural network

Yujie Bai and Min Li contributed equally to this work and should be considered co-first authors.

✉ Xiaoyi Lv
xjuwawj01@163.com

✉ Hongtao Li
lht4656@163.com

Extended author information available on the last page of the article

1 Introduction

Breast cancer is deadly cancer caused by abnormal cells that cannot be controlled [8, 9, 39]. By 2020, 2.3 million new breast cancer cases have been diagnosed, posing a serious threat to the health of women around the world [37, 42]. Therefore, early screening for breast disease is particularly necessary and significant. At present, there are many methods for early screening of breast disease, such as mammography, computed tomography technology, photoacoustic imaging, etc. [4, 14, 24, 49]. Mammography is recognized as the most effective method for the early detection of breast cancer [31]. It is primarily used for breast cancer detection [18, 43], microcalcification treatment [3], and image classification [51].

Boumaraf et al. [9] used the Breast Imaging Reporting and Data System (BI-RADS) to analyze and classify mammograms. Accurate grading results of mammograms can not only reduce mortality but also protect patients' physical and mental health by avoiding unnecessary biopsy and clinical surgery [50]. As shown in Table 1, the fifth edition of the BI-RADS classification divides breast cancer into the following 7 categories [11, 33]. However, the contrast of the mammography images of the different BI-RADS itself is very low. This may cause doctors to be easily limited by their own experience when diagnosing with the naked eye, or misdiagnosis and missed diagnosis caused by visual fatigue. Even as an expert, it is an arduous task to assign a BI-RADS category for each mammogram [6]. Therefore, this is an important research value for further studies on the classification of mammography images of breast diseases.

In recent years, there are still a host of scholars studying computer-aided diagnosis (CAD) to help doctors diagnose breast diseases. An effective CAD method can provide radiologists with reliable support for detecting abnormalities in mammograms [10, 13, 45]. Therefore, for the early diagnosis of breast diseases based on mammograms, this paper explored a computer-aided breast cancer screening and classification method based on feature extraction from a pre-trained neural network model. We explored and created a new method of image enhancement, called INCIE (Image Negatives and Contrast Limited Adaptive Histogram Equalization Image Enhancement) by classifying normal breast (BI-RADS 1), benign breast (BI-RADS 2), and probably benign breast (BI-RADS 3). The algorithm performed an image negatives operation on the dataset, and then performed a contrast-limited adaptive histogram equalization operation. After expanding the training set after INCIE, the pre-trained ResNet-50 neural network was used as the feature extractor to compare the classification results on the K-nearest neighbor (KNN)

Table 1 Fifth edition BI-RADS classification

BI-RADS	Standard
BI-RADS 0	Incomplete. Need for an additional imaging evaluation
BI-RADS 1	Normal. No abnormality
BI-RADS 2	Benign. Normal interval follow-up
BI-RADS 3	Probably benign. A short interval follow-up is recommended: 4 months follow-up for masses and 6months follow-up for micro-calcification
BI-RADS 4	Suspicious abnormality. A biopsy should be considered
BI-RADS 5	Highly suggestive of malignancy. Biopsy or surgery should be performed
BI-RADS 6	Histologically proven malignancy. Imaging is performed for cancer staging or evaluation after chemotherapy

classifier, random forest (RF) classifier, and support vector machine (SVM) classifier. The final KNN classification accuracy is 85%. This paper contributes as follows:

- Firstly, this study focuses for the first time on the subtle differences between normal, benign, and probably benign breasts so that patients can effectively avoid unnecessary biopsies. At the same time, the different classification effects of machine learning algorithms in breast disease classification were discussed, indicating that machine learning algorithms still have great potential for development in mammography diagnosis.
- Secondly, in the study, we explore for the first time the study of multiple mammogram preprocessing methods and create a new image enhancement method, which has certain reference significance.
- Finally, this study has certain application value. It can not only help radiologists to provide accurate diagnosis and effective interpretation, but also better help clinicians to provide rapid and efficient analysis results, thus reducing the workload of doctors.

2 Related works

With the deepening of relevant research, the diagnosis and grading of breast diseases have attracted the attention of some research institutions. Kumar et al. [19] analyzed machine learning algorithms such as Naïve Bayesian, Support Vector Machined, Decision Trees, etc. under the WEKA environment. Finally, the SVM algorithm achieved an accuracy of 97.89%. There are still some problems in the above research, such as incomplete experimental subjects. Many research institutions have also conducted related studies on the BI-RADS categories of breast cancer based on mammograms. Chokri et al. [11] recommended CAD to automatically classify mammograms into two categories, or into four BI-RADS categories. In this study, they used 480 images from the DDSM database for model evaluation, achieving accuracy rates of 88.02% and 83.85%. Although the study focused on the importance of different BI-RADS categories of breast cancer, there are still incomplete experimental subjects, and the difference between normal breast and other different stages has not been paid attention to. Miranda et al. [32] invited professional doctors to evaluate 46 mammograms in the DDSM database and predict four BI-RADS categories of breast diseases through a fuzzy logic inference system. The final accuracy rate was 76.67%. This study has not yet paid attention to the classification significance of normal breast, and the classification effect is poor, so we still need further study. Loizidou et al. [26] collected 100 pairs of mammograms for the classification of benign tissue (BI-RADS 1, BI-RADS 2) and suspicious tissue (BI-RADS 4, BI-RADS 5). They extracted 96 features and evaluated the performance by measuring sensitivity, specificity, accuracy, etc., in 9 different classifiers. The final result had an accuracy of 90.3% and an AUC of 0.87. However, there are still too few classification types and unreasonable divisions in this study. Zhang et al. [54] evaluated the ultrasound images of 1311 lesions according to the BI-RADS grading scale and with reference to the pathological findings. They divided benign and malignant lesions and used sensitivity, specificity, and accuracy to compare the diagnostic effects of AI and radiologists. The AI model had a diagnostic accuracy of 77.0%, a sensitivity of 82.0%, and a specificity of 71.7%. Ghaemian et al. [15] evaluated a total of 213 breast masses, and the data were analyzed using descriptive statistical methods. The sensitivity of mammography and ultrasonography alone was 72.6% and 68.9%, respectively. The accuracy rate of combined

mammography and ultrasound was 84.9%. In summary, we have summarized the relevant research papers published in recent years, and the results are shown in Table 2.

3 Materials and methods

3.1 Experimental datasets

This study selected 134 (2560*3328 Pixels, 3328*4096 Pixels) different BI-RADS mammograms of 68 breasts from January 2016 to December 2020 in the Affiliated Tumor Hospital of Xinjiang Medical University (details are shown in Table 3). Among them, BI-RADS 1 includes 25 breasts, a total of 50 mammograms, all of which are normal tissues; BI-RADS 2 includes 25 breasts, a total of 50 mammograms, all of which are benign tissues, including calcified fibers Adenoma and other signs, no malignant signs; BI-RADS 3 includes 18 breasts, a total of 34 mammograms, all of which are probably benign tissues, including clusters of fine dot-like calcifications and other signs, requiring short-term follow-up. Mammograms of all patients were retrieved from the pathology department of the hospital. All samples were taken in craniocaudal (CC) or medial oblique (MLO) position, as shown in Figs. 1 and 2, and 3.

3.2 Proposed method

The mammogram images themselves have low contrast, so assigning the correct BI-RADS category to each mammogram and thus avoiding unnecessary surgery can be a difficult task [6, 50]. To solve this problem, this paper proposes a computer-aided breast cancer screening and classification method based on feature extraction from a pre-trained neural network model. Created an INCIE image enhancement method, using the pre-trained ResNet-50 model as the feature extractor, input the extracted features into the KNN classifier, RF classifier, and SVM classifier respectively, and plotted the ROC curve and confusion matrix [15, 41, 54], to evaluate the experimental results. Figure 4 illustrates the framework of the early detection method for breast disease proposed in this study.

3.3 Preprocessing

The initial process of image classification is the preprocessing stage [36]. The preprocessing technology improves the accuracy of the classification by improving the quality of the image. Different researchers have different preprocessing methods for mammograms. For example, wiener filtering method used to filter equalized image sets in image processing [23, 36]; median filtering method to protect edge information [1]; laplacian filtering method for preprocessing because of its simplicity and ability to eliminate noise [5]; CLAHE for enhancing the contrast of mammograms [31]. In addition, we also discuss the contrast stretching method and the image negatives method. This paper involves removing the artifacts and background, Eliminating the pectoral muscles, and the application of image enhancement [7]. The specific steps are as follows: Our first task is to convert the CR sequence diagram in DICOM format to JPG format. Then using LabelMe software, two professional doctors help to manually outline the pectoral muscles part in the mammogram, eliminate the pectoral muscles

Table 2 Summary of research literature related to the early detection and diagnosis of breast disease

Author	Work	Year	Research task	Advantages and disadvantages
Miranda et al.	[32]	2015	BI-RADS grading of mammograms	This study explores the differences between mammography images of different grades. Prediction of BI-RADS categories for four breast diseases by a fuzzy logic inference system. The method finally achieved an accuracy of 76.67%. However, this study has not paid attention to the classification significance of normal breast, and the classification effect is poor, and further research is needed.
Chokri et al.	[11]	2016	Binary classification and quaternary classification of mammograms	In this study, they automatically extracted 22 features for prediction with an accuracy of 88.02% and 83.85%, respectively. Although this study focused on the importance of different BI-RADS classifications of breast cancer, there are still incomplete experimental subjects and the differences between normal breast and other different stages have not been appreciated. At the same time, the data pre-processing is not comprehensive enough.
Kumar et al.	[19]	2019	Binary classification of mammograms	This study analyzed the classification of benign and malignant tumors using machine learning algorithms such as Naive Bayes, Support Vector Machines, and Decision Trees. In the end, the support vector machine algorithm achieved an accuracy of 97.89%. However, there are also problems with incomplete subjects in the study, and the test results are more extensive, which may lead to patients receiving unnecessary biopsies.
Loizidou et al.	[26]	2021	Classification of benign and suspicious tissues	This study extracted 96 features and evaluated the performance by measuring sensitivity, specificity, accuracy, etc., in 9 different classifiers. However, this study did not discuss BI-RADS 2 and BI-RADS 3 in detail, resulting in an unreasonable division.
Zhang et al.	[54]	2021	BI-RADS grading of breast ultrasound images	This study divided benign and malignant lesions and used sensitivity, specificity, and accuracy to compare the diagnostic effects of AI and radiologists. The AI model had a diagnostic accuracy of 77.0%, a sensitivity of 82.0%, and a specificity of 71.7%. However, the experimental subjects in this study are relatively small, and the experimental accuracy still needs to be improved.
Ghaemian et al.	[15]	2021	Combined examination of mammograms and ultrasound images	the data were analyzed using descriptive statistical methods. The sensitivity of mammography and ultrasonography alone was 72.6% and 68.9%, respectively. The accuracy rate of combined mammography and ultrasound was 84.9%. However, there are different characteristics of different types of images and a certain lack of discussion when performing image processing, while there is room for improvement in the results of single examination.

Table 2 (continued)

Author	Work Year	Research task	Advantages and disadvantages
Our approach		BI-RADS grading of mammograms	In our study, a new image enhancement framework was created for the first time based on a comparison of a set of multiple data preprocessing methods for the detection of normal, benign, and probably benign breasts, achieving 85 on the KNN classifier % accuracy with an AUC of 0.89.

and remove the artifacts and background, and keep them intact breast part. The concrete implementation is shown in Fig. 5.

In addition, we also explored the effect of different preprocessing methods. According to the large black area in the mammography target, we need to highlight the characteristics of the white and gray parts, and we chose the image negatives method for processing. To improve the local contrast of the image, we perform contrast limited adaptive histogram equalization (CLAHE) based on the above operation and resize the image to 1024×1024 pixels [1, 31]. The formula of image negatives is as follows (1).

$$s = L - 1 - r \quad (1)$$

Among them, s is the gray value of a certain point of the inverted target image, L represents the gray level, and r represents the pixel value of the original image pixel. The concrete implementation is shown in Fig. 6.

3.4 Data augmentation

Data augmentation is an attractive solution to reduce model overfitting, improve model generalization and performance [2, 17]. Currently, the publicly available mammograms are limited. Therefore, we adopted data enhancement to prevent overfitting caused by small sample data sets [24]. Data augmentation does not only increase the amount of our data but makes the data set “stronger.” In this study, we randomly divided the dataset with a ratio of 7:3. The training set is augmented with horizontal and vertical flipping, translation by 10 pixels, rotation by 90° , 180° , and 270° , and adding noise [20, 22, 37, 41]. The augmented training set contains a total of 9400 mammography images.

Table 3 The obtained experimental images

Type	Number of patients	Number of images
Normal (BI-RADS 1)	25	50
Benign (BI-RADS 2)	25	50
Probably benign (BI-RADS3)	18	34
Total	68	134

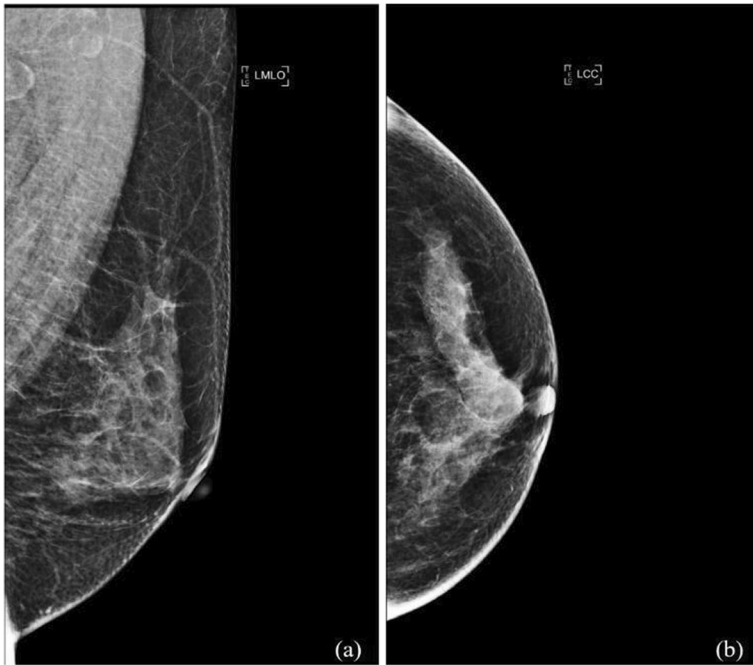


Fig. 1 Mammograms of the normal breast. **a** Mediolateral Oblique (MLO) of the normal breast (BI-RADS 1), **b** Craniocaudal (CC) of the normal breast (BI-RADS 1)

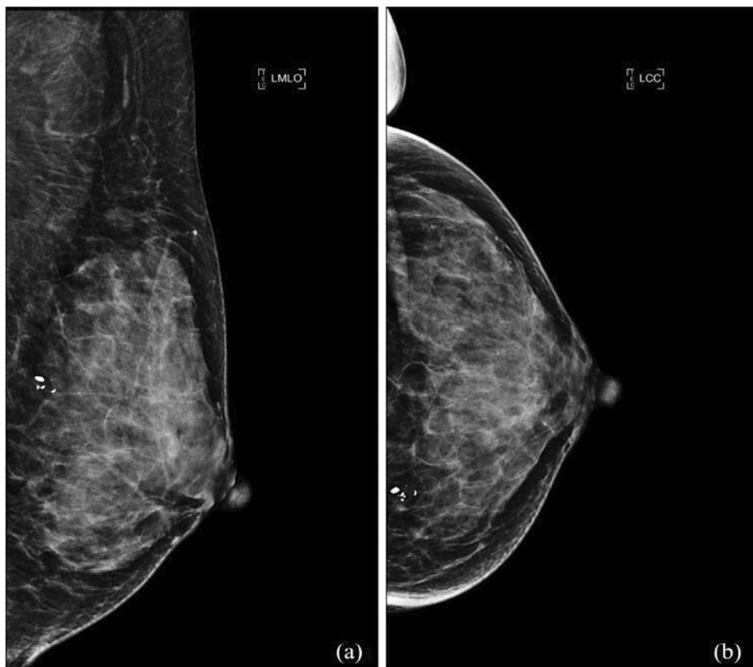


Fig. 2 Mammograms of the benign breast. **a** Mediolateral Oblique (MLO) of the benign breast (BI-RADS 2), **b** Craniocaudal (CC) of the benign breast (BI-RADS 2)

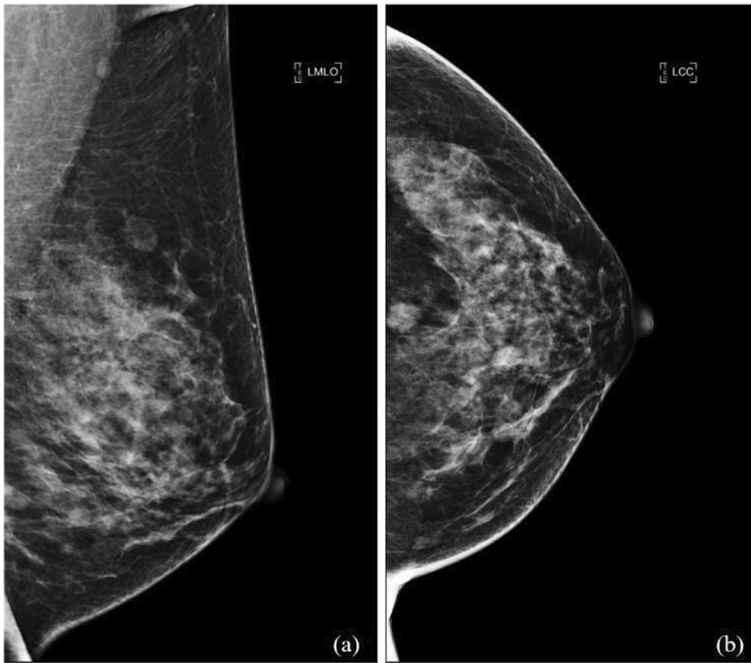


Fig. 3 Mammograms of the probably benign breast. **a** Mediolateral Oblique (MLO) of the probably benign breast (BI-RADS 3), **b** Craniocaudal (CC) of the probably benign breast (BI-RADS 3)

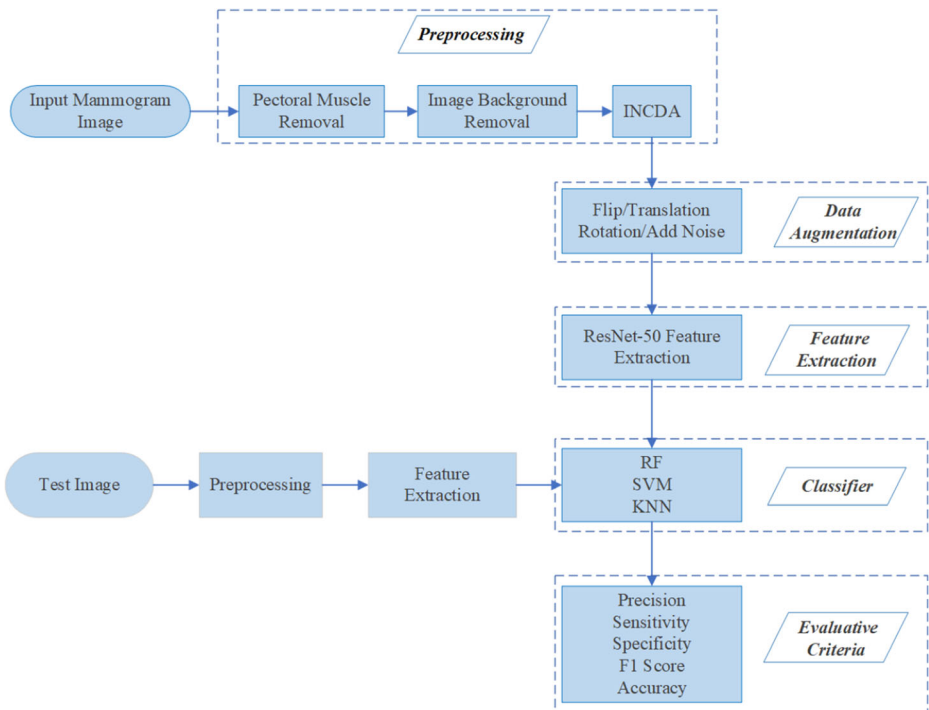


Fig. 4 Overall process of proposed model. The blue boxes in the figure represent the five main parts in this paper

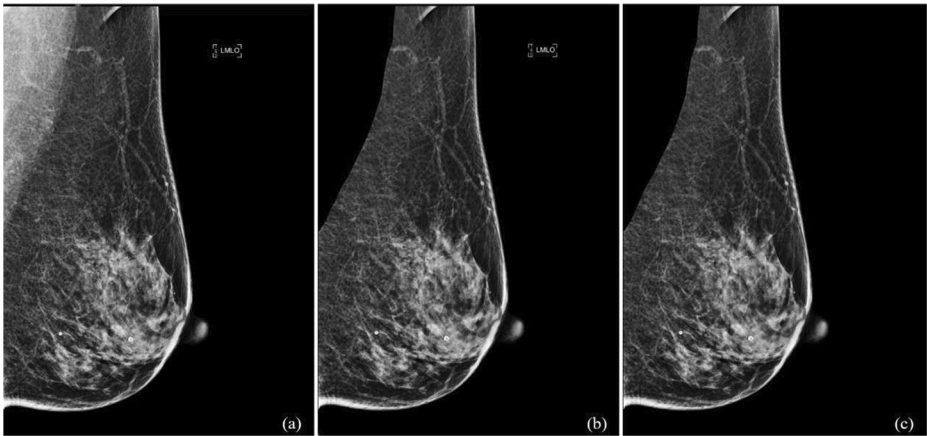


Fig. 5 Eliminate the pectoral muscles and remove the artifacts and background. **a** Manually outline the area, **b** Eliminate the pectoral muscle, **c** Remove the artifacts and background

3.5 Feature extraction

According to research, traditional classification algorithms consume a lot of time and energy when deciding which feature extraction algorithm to use. However, CNN is widely used because of its special structure of local weight sharing, which has unique advantages in the field of image processing [16, 21, 40]. The ResNet was proposed by K He et al. in 2015, and it has attracted much attention due to its excellent performance in image classification and target detection [55]. At present, researchers have applied ResNet-50 for the feature extraction of mammograms and have achieved good classification results [27, 30, 35]. Although Mammograms are slightly different from natural image data, they can still detect basic features such as

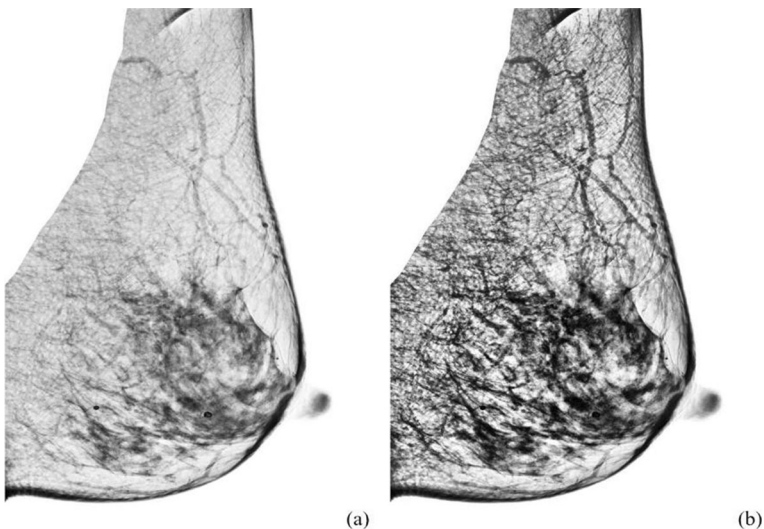


Fig. 6 Image Negatives and Contrast limited adaptive histogram equalization Image Enhancement (INCIE) **a** Image Negatives and Contrast limited adaptive histogram equalization Image Enhancement (INCIE), **b** Image Negatives and Contrast limited adaptive histogram equalization Image Enhancement (INCIE).

edges or shapes through pre-trained neural network models [16, 52]. Considering the complexity of the feature extraction and classification process, this research uses the ResNet-50 model to extract features from the image.

3.6 Classifier

3.6.1 RF classifier

In 2001, Breiman modified bagging and proposed a random forest algorithm [29]. The essence of the algorithm is an improvement of the decision tree algorithm. It is flexible and one of the most practiced integrated classifiers, which has proven its efficiency and superiority in many classification applications [46]. A test sample can select the most probable classification after counting the classification results of each tree [12, 34]. The RF implementation is relatively simple and suitable for parallel computing. Each tree randomly selects samples and randomly selects features [38, 47]. Since the number of trees ($n_{\text{estimator}}$) directly affects the final classification effect, during the experiment, we had to determine how many trees to take.

3.6.2 SVM classifier

SVM is a supervised learning algorithm proposed by Vapnik et al. [44], which is mainly used to solve data classification in the field of pattern recognition. The biggest advantage of this algorithm is that it's backed up by rigorous mathematical theory. In the vector space composed of sample points, SVM achieves the effect of data classification and prediction by finding a partitioning hyperplane that can correctly separate the two types of data on both sides [25, 53]. SVM classifies samples of different categories by dividing the hyperplane with the "maximum interval" suitable for sample classification. As shown in Formula (2):

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m l_i(y(w^T x_i + b) - 1) \cdot l_i \quad (2)$$

Where l_i is the loss function and C is the penalty coefficient. This paper uses the Lagrangian multiplier method to solve the problem. The equivalent conversion of the dual problem is given in Eq. (3).

$$\max_{\alpha} = \sum_i^m \alpha_i - \frac{1}{2} \sum_i^m \sum_j^m \alpha_i \alpha_j y_i y_j x_i^T x_j \cdot s.t. \sum_i^m \alpha_i y_i = 0, 0 \leq \alpha_i \leq C \quad (3)$$

Equation (3) shows that selecting an appropriate kernel function is very important because it not only can reduce the calculation of the inner product of $x_i^T x_j$ but can also improve the accuracy of classification.

3.6.3 KNN classifier

KNN algorithm can be used not only for regression but also for classification, a statistical method for pattern recognition, which plays a significant role in the field of machine learning [19]. The basic thought of the KNN algorithm is just like one in real life "Birds of a feather

flock together.“ First, calculate the distance between the image feature to be classified and the training image feature, and then sort. Then determine a new sample category based on the category of similar training data features [48]. For a data sample X to be classified, its K-nearest neighbors are searched and then X is assigned to the class label to which most of its neighbors belong. The choice of K also affects the performance of the K-nearest neighbor algorithm [28]. It is concise and clear, easy to implement, has fast model training time, good prediction effect, and is especially suitable for multi-classification problems. The KNN classifier works as follows:

1. Initialize the value of K.
2. Calculate the distance between the input sample and training samples.
3. Sort the distances.
4. Take top K-nearest neighbors.
5. Apply simple majority.
6. Predict class labels with more neighbors for the input sample.

3.7 Evaluative criteria

In this experiment, we conduct model performance evaluation by drawing the ROC curve and confusion. When testing the classification model of this study, we calculated the probabilities of various test samples. We can deduce the true-positive rates (TPR) and false-positive rates (FPR) and then use these TPR-FPR data to draw the ROC curve and calculate the area AUC under the ROC curve. Another name for the confusion matrix is the error matrix. It is a master format for showing precision evaluation. These precision indexes reflect the precision of mammograms classification from diverse sides. In the confusion matrix, all the correct prediction results are on the diagonal, so we can easily and intuitively see the error from it. In this experiment, we use the following indicators to evaluate the parameters: precision, sensitivity, specificity, and F1 score. The formula is (4)–(7).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

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

$$Specificity = \frac{TN}{FP + TN} \quad (6)$$

$$F1 \text{ Score} = 2 \cdot \frac{Precision \cdot Sensitivity}{Precision + Sensitivity} \quad (7)$$

Table 4 Experimental environment

Category	Name	Version
CPU	Intel(R)Core(TM) i5-7200U	--
RAM	4.00 GB	--
GPU	NVIDIA GeForce 940MX	--
Image processing	Matlab	2016a
Development language	Python	V3.6.2
Data processing library	Numpy	V1.19.5
Machine learning library	Sklearn	V0.24.2
Deep learning framework 1	TensorFlow	V2.0.0
Deep learning framework 2	Keras	V2.2.0

4 Results and discussion

4.1 Experimental setup

In this experiment, the Matlab2016a platform is used to achieve data preprocessing, the LabelMe tool is used to outline the breast pectoral muscles, and the convolutional neural network is implemented using TensorFlow. All experiments are performed on an Intel(R) Core (TM) i5-7200U CPU @ 2.50 GHz 2.71 GHz machine. The software and hardware conditions of the specific experimental environment are shown in Table 4.

4.2 Experimental results

4.2.1 Preprocessing results

In this section, the classification results produced by different preprocessing methods are analyzed in detail. Table 5 shows the accuracy of classification on different classifiers using different preprocessing methods.

Table 5 Comparison of results of different image preprocessing methods

Methods		RF	SVM	KNN
Median Filter	3×3	45.0%	52.5%	50.0%
	5×5	47.5%	50.0%	47.5%
	7×7	47.5%	60.0%	47.5%
	11×11	55.0%	50.0%	55.0%
Wiener Filter	3×3	55.0%	57.5%	47.5%
	5×5	40.0%	42.5%	45.0%
	7×7	55.0%	65.0%	50.0%
	11×11	50.0%	52.5%	55.0%
Laplacian Filter	−4	60.0%	47.5%	40.0
	−8	60.0%	57.5%	45.0%
CLAHE		47.5%	52.5%	42.5%
Contrast Stretch		42.5%	55.0%	37.5%
Image Negatives		50.0%	47.5%	52.5%
Image Negatives	CLAHE	60.0%	60.0%	57.5%

To find a suitable preprocessing method that is helpful for accurate classification, this experiment uses a variety of methods for comparison experiments [1, 5, 31, 36]. During the experiment, we considered that different filter windows and the size of the diagonal element center would produce images with different effects. Therefore, when we experiment with median filtering and Wiener filtering methods, we choose 3×3 , 5×5 , 7×7 , and 11×11 windows for comparison; When using the Laplacian filtering method to experiment, the diagonal element center is selected as -4 and -8 for comparison. Most importantly, we have processed the mammogram using image negatives technology based on its characteristics. Experimentally, the INCIE method we created proved to be the most effective and better able to show the feature details in the molybdenum target images at different levels.

4.2.2 Data augmentation results

Since there are few publicly available mammograms at present, we use data augmentation to expand the number of images in an attempt to increase the generalization performance [24]. According to the above preprocessing experiment, we selected the four groups of data with the best performance in Table 5 for data augmentation. Data augmentation methods are translation, rotation, and flip. As shown in Table 6, the data on the left of each model is the result of preprocessing, and the right data is the result of data augmentation. The experiments show that the pre-processed images can be enhanced by data, which can make the classification effect better. In particular, the image processed by image negatives and CLAHE have the largest improvement in the KNN classifier, and the accuracy has increased by 25%.

In addition, we found that many researchers also pay attention to increasing noise data when performing data augmentation for mammograms [54]. Therefore, we chose the best-performing data augmentation method in Table 6 to add noise processing, and the classification effect on the KNN classifier increased by 2.5% (as shown in Table 7). The results show that when classifying mammograms, adding noise data has a certain effect on improving the performance of the model.

4.2.3 Comparison of KNN classifier with others

In this study, the KNN classifier was used to classify mammograms, and the current two mainstream classification algorithms SVM and RF were selected to compare precision, sensitivity, specificity, F1-score, and accuracy. The comparison results are shown in Table 6. According to Table 6, we can know that when facing BI-RADS 1, the sensitivity of the KNN classifier is as high as 100.0%, which is 13.3% higher than the RF classifier and 20.0% higher than the SVM classifier. For BI-RADS 2, the precision of the KNN classifier

Table 6 Comparison of results before and after data enhancement

Methods	RF		SVM		KNN	
Median Filter 11×11	55.0%	60.0%	50.0%	70.0%	55.0%	52.5%
Wiener Filter 7×7	55.0%	65.0%	65.0%	62.5%	50.0%	60.0%
Laplacian Filter -8	60.0%	65.0%	57.5%	57.5%	45.0%	65.0%
Image Negatives CLAHE	60.0%	62.5%	60.0%	72.5%	57.5%	82.5%

Table 7 Comparison of results of different data augmentation methods

Methods	RF	SVM	KNN
Translation / Rotate / Flip	62.5%	72.5%	82.5%
Translation / Rotate / Flip / Add Noise	62.5%	67.5%	85.0%

reaches 91.7%, which is 35.4% and 29.2% higher than the RF classifier and SVM classifier severally. In BI-RADS 3, the F1-score of the KNN classifier is 84.2%. However, the results of the RF classifier and SVM classifier are not very satisfactory, which are 42.9% and 62.5%, respectively. The results show that all the evaluation criteria perform better on the KNN classifier. The accuracy of the KNN reaches 85.0%, which proves the rationality of choosing the KNN classifier (Table 8).

4.2.4 ROC curve and confusion matrix

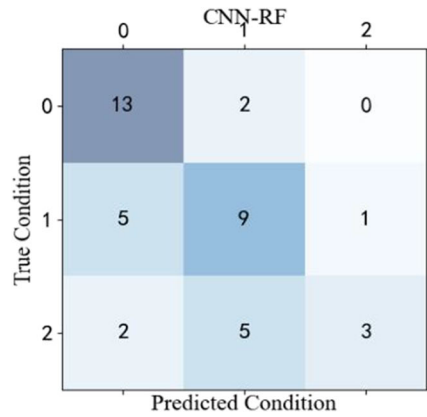
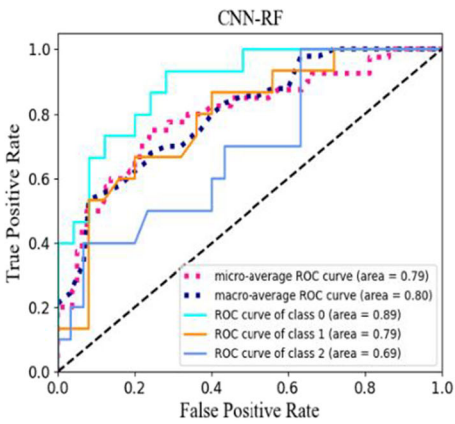
In addition, we also perform performance evaluation by drawing ROC curves and confusion matrices. As shown in Fig. 7, (a), (b), and (c) represent the ROC curves and confusion matrices of the classification results of RF, SVM, and KNN severally ('0' means BI-RADS 1, and '1' means BI-RADS 2, '2' means BI-RADS 3). According to Fig. 7, we can see that the micro-average AUC value of the KNN classifier reaches 0.89, which is 0.1 and 0.02 higher than the RF classifier and the SVM classifier, respectively. The AUC values of the KNN classifier under BI-RADS 1, BI-RADS 2, and BI-RADS 3 are 0.92, 0.85, and 0.88, respectively. Compared with the RF classifier and the SVM classifier, when the KNN classifier is adopted, the AUC value of BI-RADS 3 has the most significant increase, increasing by 0.19 and 0.05 severally. The results show that KNN has a higher recognition rate for mammograms of three BI-RADS categories, which more effectively proves that the KNN classifier is our best choice.

4.3 Discussion

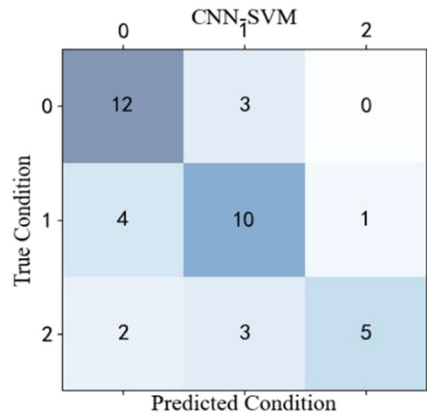
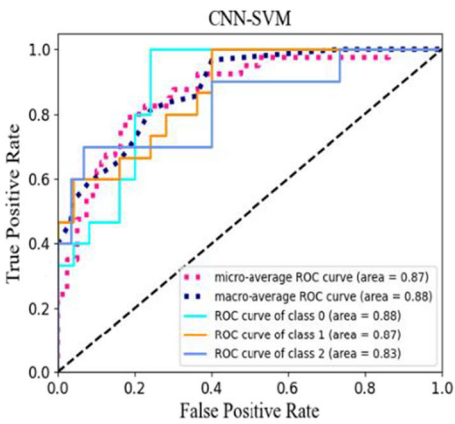
In this paper, we propose an INCIE image enhancement method to process mammogram images with respect to their own characteristics. Experiments show that our proposed method provides more detailed feature information, which lays a good foundation for later classification. Through the horizontal comparison of the three classifiers and

Table 8 Performance analysis of RF, SVM and KNN models

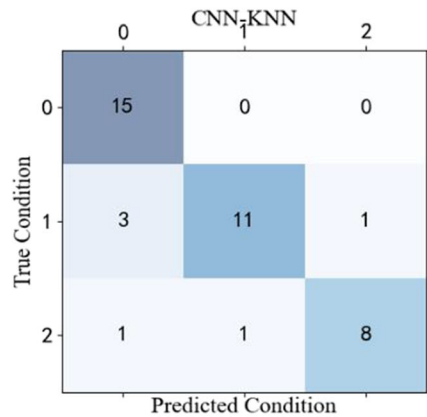
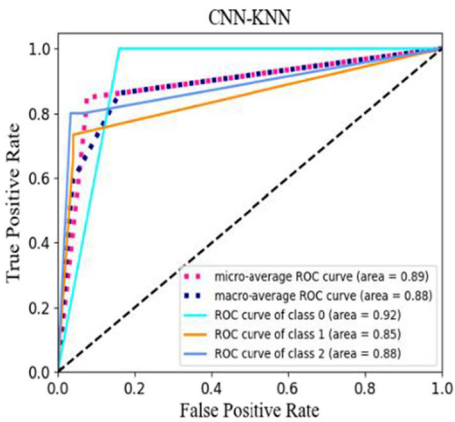
Model	Type	Precision	Sensitivity	Specificity	F1-score	Accuracy
RF	BI-RADS 1	65.0%	86.7%	63.2%	74.3%	62.5%
	BI-RADS 2	56.3%	60.0%	69.6%	58.1%	
	BI-RADS 3	75.0%	30.0%	95.7%	42.9%	
SVM	BI-RADS 1	66.7%	80.0%	71.4%	72.7%	67.5%
	BI-RADS 2	62.5%	66.7%	65.4%	64.5%	
	BI-RADS 3	83.3%	50.0%	95.7%	62.5%	
KNN	BI-RADS 1	78.9%	100.0%	82.6%	88.2%	85.0%
	BI-RADS 2	91.7%	73.3%	95.8%	81.5%	
	BI-RADS 3	88.9%	80.0%	96.3%	84.2%	



(a)



(b)



(c)

Fig. 7 ROC curves and confusion matrices. **a** ROC curves and confusion matrices of RF classifier, **b** ROC curves and confusion matrices of SVM classifier, **c** ROC curves and confusion matrices of KNN classifier

different evaluation metrics, it is found that RF classifiers do not achieve ideal results at different levels of classification. We believe that the classification performance is poor due to the small sample size. When the RF classifier is applied to mammography images with high feature similarity, the classification results of BI-RADS 3 are more inclined to the category with a large number of samples. The evaluation metrics of the SVM classifier are slightly better than the RF classifier. Although the classification ability of SVM makes up for the lack of classification ability of small sample datasets in the field of deep learning to a certain extent, the classification results of BI-RADS 3 under small samples still have not achieved satisfactory results. Therefore, we choose the One-vs-the-rest (OVR) multiclass strategy for discussion, and the experimental results are shown in Fig. 8. The experimental results show that the classification accuracy of the OVR-SVM model is improved by 2.5%, and it is superior to the SVM model in the classification accuracy and performance of different types of breast diseases. The KNN classifier achieves better results in all three breast classes, which shows that the KNN classifier is suitable for multi-classification problems and has good classification performance and generalization ability for datasets with more similar features in mammogram images. At the same time, it helps radiologists to provide accurate diagnosis and effective interpretation to a certain extent and helps clinicians to provide fast and efficient analysis results, thus reducing the workload of doctors.

4.4 Study limitations

This study has some limitations. First, since the data in this study were collected from local hospitals, there are certain differences in the datasets annotated by different radiologists, so the data generality is not strong. Second, the number of mammography images, included in this study was not large enough and the types were not comprehensive enough to warrant that the conclusions of this study could be justified by other images. Finally, the mammograms were scaled down to fit the available GPU, and future studies can maintain the resolution of the original images, provide finer detail features, and potentially improve performance.

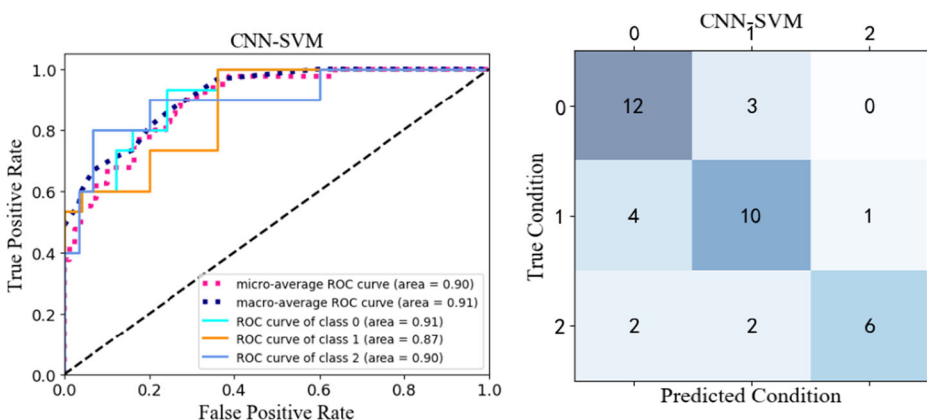


Fig. 8 ROC curves and confusion matrices of SVM classifier using OVR strategy

5 Conclusion

One of the best early detection approaches for breast cancer is a classification based on mammography images. It not only enables patients to receive more appropriate treatment options but also effectively avoids unnecessary surgery. In this study, intelligent classification was performed for early mammogram images of different grades. Since mammogram images are characterized by predominantly black color, we created a new image enhancement framework INCIE based on a set of multiple data preprocessing methods compared for the first time. It enhances the white or gray areas in the image to better show subtle differences between different categories of mammograms. In the experiments, considering the complexity of feature extraction and classification, we adopted the pre-trained ResNet-50 neural network model for feature extraction, which was evaluated in three different classifiers. The results showed that the KNN classifier has the best classification effect, with an accuracy rate of 85% and an AUC of 0.89. This method is practical and reliable in the diagnosis of early breast disease.

As an applied study, this study has a certain auxiliary effect on the diagnosis of physicians, while solving the time-consuming and laborious problems associated with traditional manual film reading. In future work, we will continue to collect more types of data from different central institutions to build better-performing and more generalizable auxiliary diagnostic models. At the same time, we will combine multi-directional mammograms for analysis to maximize the significance of the early diagnosis of breast diseases (Appendix Table 9 and 10).

Appendix 1

Table 9 List of abbreviations

Serial Number	Full Name	Abbreviations
1	Breast Imaging Reporting and Data System	BI-RADS
2	Computer-aided diagnosis	CAD
3	Normal Breast	BI-RADS 1
4	Benign breast	BI-RADS 2
5	Probably benign breast	BI-RADS 3
6	Image Negatives and Contrast Limited Adaptive Histogram Equalization Image Enhancement	INCIE
7	K-nearest neighbor	KNN
8	Random forest	RF
9	Support vector machine	SVM
10	Craniocaudal	CC
11	Medial oblique	MLO
12	Contrast limited adaptive histogram equalization	CLAHE
13	True-positive rates	TPR
14	False-positive rates	FPR
15	One-vs-the-rest	OVR

Appendix 2

Table 10 List of symbols

Serial Number	Symbols	Meaning
1	s	The gray value of a certain point of the inverted target image
2	L	The gray level
3	r	The pixel value of the original image pixel
4	$n_{\text{-estimator}}$	The number of trees
5	l_i	The loss function
6	C	The penalty coefficient
7	TP	True positive
8	TN	True negative
9	FP	False positive
10	FN	False negative

Acknowledgements This work was supported in part by the Xinjiang Uygur Autonomous Region Science Foundation for Distinguished Young Scholars under Grant 2019Q003, in part by the Tianshan Innovation Team Planning Project under Grant 2020D14031, and in part by the Tianshan Youth Planning Project under Grant 2019Q043.

Declarations

Conflict of interest The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose.

References

1. Abdel-Nasser M, Moreno A, Puig D (2016) Temporal mammogram image registration using optimized curvilinear coordinates. *Comput Methods Programs Biomed* 127:1–14
2. Abdelhafiz D et al (2019) Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinform* 20(Suppl 11):281
3. Al-Najdawi N, Biltawi M, Tedmori S (2015) Mammogram image visual enhancement, mass segmentation and classification. *Appl Soft Comput* 35:175–185
4. Alshayegi MH et al (2022) Computer-aided detection of breast cancer on the Wisconsin dataset: an artificial neural networks approach. *Biomed Signal Process Control* 71:103141
5. Arputham C et al (2021) Mammographic image classification using deep neural network for computer-aided diagnosis. *Intell Autom Soft Comput* 27(3):747–759
6. Bakkouri I, Afdel K (2018) Multi-scale CNN based on region proposals for efficient breast abnormality recognition. *Multimed Tools Appl* 78(10):12939–12960
7. Beeravolu AR et al (2021) Preprocessing of breast cancer images to create datasets for Deep-CNN. *IEEE Access* 9:33438–33463
8. Benhassine NE, Boukaache A, Boudjehem D (2019) Classification of mammogram images using the energy probability in frequency domain and most discriminative power coefficients. *Int J Imaging Syst Technol* 30(1):45–56
9. Boumaraf S et al (2020) A new computer-aided diagnosis system with modified genetic feature selection for BI-RADS classification of breast masses in mammograms. *Biomed Res Int* 2020:7695207
10. Bozkurt S et al (2016) Using automatically extracted information from mammography reports for decision-support. *J Biomed Inform* 62:224–231
11. Chokri F, Hayet M, Farida (2016) Mammographic mass classification according to Bi-RADS lexicon. *IET Comput Vision* 11(3):189–198
12. Corcoran J et al (2015) The effects of point or polygon based training data on randomforest classification accuracy of wetlands. *Remote Sens* 7(4):4002–4025

13. Fanizzi A et al (2016) Automatised detection of microcalcification in mammography. *Phys Med* 32:217
14. Ferlay J et al (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 127(12):2893–2917
15. Ghaemian N, Haji Ghazi N, Tehrani, Nabahati M (2021) Accuracy of mammography and ultrasonography and their BI-RADS in detection of breast malignancy. *Casp J Intern Med* 12(4):573–579
16. Goel N et al (2022) Dilated CNN for abnormality detection in wireless capsule endoscopy images. *Soft Comput* 26(3):1231–1247
17. Goel N et al (2022) Investigating the significance of color space for abnormality detection in wireless capsule endoscopy images. *Biomed Signal Process Control* 75:103624
18. Greenspan H, van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 35(5):1153–1159
19. Kumar A, Sushil R, Tiwari AK (2019) Comparative study of classification techniques for breast cancer diagnosis. *Int J Comput Sci Eng* 7(1):234–240
20. Lbakhir IA et al (2017) A New mammogram preprocessing method for computer-aided diagnosis systems. In: 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), pp 166–171
21. Lecun Y et al (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
22. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
23. Lee S et al (2019) Noise removal in medical mammography images using fast non-local means denoising algorithm for early breast cancer detection: a phantom study. *Optik* 180:569–575
24. Li H et al (2019) Benign and malignant classification of mammogram images based on deep learning. *Biomed Signal Process Control* 51:347–354
25. Li M et al (2020) Computer-aided diagnosis and staging of pancreatic cancer based on CT images. *IEEE Access* 8:141705–141718
26. Loizidou K et al (2021) Digital subtraction of temporally sequential mammograms for improved detection and classification of microcalcifications. *Eur Radiol Exp* 5(1):40
27. Lu L, Daigle BJ Jr (2020) Prognostic analysis of histopathological images using pre-trained convolutional neural networks: application to hepatocellular carcinoma. *PeerJ* 8:e8668
28. Luque-Baena RM et al (2014) Application of genetic algorithms and constructive neural networks for the analysis of microarray cancer data. *Theor Biol Med Model* 11:57
29. Maroof N et al (2020) Mitosis detection in breast cancer histopathology images using hybrid feature space. *Photodiagnosis Photodyn Ther* 31:101885
30. Mathur M et al (2020) Crosspooled FishNet: transfer learning based fish species classification model. *Multimed Tools Appl* 79(41–42):31625–31643
31. Mehmood M et al (2021) Machine learning enabled early detection of breast cancer by structural analysis of mammograms. *Comput Mater Contin* 67(1):641–657
32. Miranda GH, Felipe JC (2015) Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization. *Comput Biol Med* 64:334–346
33. Nwadike UI et al (2017) Mammographic classification of breast lesions amongst women in Enugu, South East Nigeria. *Afr Health Sci* 17(4):1044–1050
34. Pavay TG et al (2017) Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport* 20(1):75–80
35. Rajaraman S et al (2018) Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ* 6:e4568
36. Rajathi GM (2020) Optimized radial basis neural network for classification of breast cancer images. *J Ambient Intell Humaniz Comput*
37. Rehman KU et al (2021) Computer vision-based microcalcification detection in Digital Mammograms using fully connected depthwise separable convolutional neural network. *Sensors (Basel)* 21(14):4854
38. Rigatti SJ (2017) Random forest. *J Insurance Med (New York, NY)* 47(1):31–39
39. Saffari N et al (2020) Fully automated breast density segmentation and classification using deep learning. *Diagnostics (Basel)* 10(11):988
40. Sampath D, Murthy A, Karthikeyan T, Vinoth R, Kanna (2021) Gait-based person fall prediction using deep learning approach. *Soft Comput*
41. Shen L et al (2019) Deep learning to improve breast cancer detection on screening mammography. *Sci Rep* 9(1):12495
42. Sung H 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(3):209–249
43. Tajbakhsh N et al (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 35(5):1299–1312

44. Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10(5):988–999
45. Verma B (2008) Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms. *Artif Intell Med* 42(1):67–79
46. Williamson S, Vijayakumar K, Kadam VJ (2021) Predicting breast cancer biopsy outcomes from BI-RADS findings using random forests with chi-square and MI features. *Multimed Tools Appl*
47. Xue J, Zhao YX (2008) Random forests of phonetic decision trees for acoustic modeling in conversational speech recognition. *IEEE Trans Audio Speech Lang Process* 16(3):519–528
48. Yan Z et al (2020) Rapid identification of benign and malignant pancreatic tumors using serum Raman spectroscopy combined with classification algorithms. *Optik* 208:164473
49. Yang B et al (2021) Detection of breast cancer of various clinical stages based on serum FT-IR spectroscopy combined with multiple algorithms. *Photodiagnosis Photodyn Ther* 33:102199
50. Zeng J et al (2019) A probabilistic model to support radiologists' classification decisions in mammography practice. *Med Decis Making* 39(3):208–216
51. Zhang X et al (2015) Towards large-scale histopathological image analysis: hashing-based image retrieval. *IEEE Trans Med Imaging* 34(2):496–506
52. Zhang X et al (2017) Whole mammogram image classification with convolutional neural networks. In: Hu XH et al (Eds) 2017 IEEE International Conference on Bioinformatics and Biomedicine, pp 700–704
53. Zhang Q, Wang H, Yoon SW (2020) A 1-norm regularized linear programming nonparallel hyperplane support vector machine for binary classification problems. *Neurocomputing* 376:141–152
54. Zhang N et al (2021) Application of deep learning to establish a diagnostic model of breast lesions using two-dimensional grayscale ultrasound imaging. *Clin Imaging* 79:56–63
55. Zhang S et al (2021) Research on application of classification model based on Stack generalization in staging of cervical tissue pathological images. *IEEE Access* 9:48980–48991

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Affiliations

Yujie Bai¹ · Min Li² · Xiaojian Ma³ · Xiaojing Gan³ · Cheng Chen¹ · Chen Chen^{2,5} · Xiaoyi Lv^{1,4,5,6} · Hongtao Li³

¹ College of Software, Xinjiang University, Urumqi 830046, China

² College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

³ The Affiliated Cancer Hospital of Xinjiang Medical University, Urumqi 830011, China

⁴ Key Laboratory of Software Engineering Technology, Xinjiang University, Urumqi 830046, China

⁵ Department of Pathology, Xinjiang Key Laboratory of Clinical Genetic Testing and Biomedical Information, Karamay Central Hospital, Karamay, China

⁶ Key Laboratory of Signal Detection and Processing, Xinjiang University, Urumqi 830046, China