Am J Transl Res 2024;16(6):2411-2422
www.ajtr.org /ISSN:1943-8141/AJTR0156732

Original Article

Deep learning modeling using mammography images for predicting estrogen receptor status in breast cancer

Wenfeng Duan[1], Zhiheng Wu[2], Huijun Zhu[2], Zhiyun Zhu[3], Xiang Liu[1], Yongqiang Shu[1], Xishun Zhu[4], Jianhua Wu[2], Dechang Peng[1]

[1]Department of Radiology, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China; [2]School of Information Engineering, Nanchang University, Nanchang, Jiangxi, China; [3]Department of Cardiology, Jiangxi Provincial People's Hospital, Nanchang, Jiangxi, China; [4]School of Advanced Manufacturing, Nanchang University, Nanchang, Jiangxi, China

Received March 21, 2024; Accepted May 12, 2024; Epub June 15, 2024; Published June 30, 2024

Abstract: Background: The estrogen receptor (ER) serves as a pivotal indicator for assessing endocrine therapy efficacy and breast cancer prognosis. Invasive biopsy is a conventional approach for appraising ER expression levels, but it bears disadvantages due to tumor heterogeneity. To address the issue, a deep learning model leveraging mammography images was developed in this study for accurate evaluation of ER status in patients with breast cancer. Objectives: To predict the ER status in breast cancer patients with a newly developed deep learning model leveraging mammography images. Materials and methods: Datasets comprising preoperative mammography images, ER expression levels, and clinical data spanning from October 2016 to October 2021 were retrospectively collected from 358 patients diagnosed with invasive ductal carcinoma. Following collection, these datasets were divided into a training dataset (n = 257) and a testing dataset (n = 101). Subsequently, a deep learning prediction model, referred to as IP-SE-DResNet model, was developed utilizing two deep residual networks along with the Squeeze-and-Excitation attention mechanism. This model was tailored to forecast the ER status in breast cancer patients utilizing mammography images from both craniocaudal view and mediolateral oblique view. Performance measurements including prediction accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curves (AUCs) were employed to assess the effectiveness of the model. Results: In the training dataset, the AUCs for the IP-SE-DResNet model utilizing mammography images from the craniocaudal view, mediolateral oblique view, and the combined images from both views, were 0.849 (95% CIs: 0.809-0.868), 0.858 (95% CIs: 0.813-0.872), and 0.895 (95% CIs: 0.866-0.913), respectively. Correspondingly, the AUCs for these three image categories in the testing dataset were 0.835 (95% CIs: 0.790-0.887), 0.746 (95% CIs: 0.793-0.889), and 0.886 (95% CIs: 0.809-0.934), respectively. A comprehensive comparison between performance measurements underscored a substantial enhancement achieved by the proposed IP-SE-DResNet model in contrast to a traditional radiomics model employing the naive Bayesian classifier. For the latter, the AUCs stood at only 0.614 (95% CIs: 0.594-0.638) in the training dataset and 0.613 (95% CIs: 0.587-0.654) in the testing dataset, both utilizing a combination of mammography images from the craniocaudal and mediolateral oblique views. Conclusions: The proposed IP-SE-DResNet model presents a potent and non-invasive approach for predicting ER status in breast cancer patients, potentially enhancing the efficiency and diagnostic precision of radiologists.

Keywords: Deep learning, breast cancer, estrogen receptor, mammography, radiomics

Introduction

Breast cancer stands as the most frequently diagnosed cancer in females, holding the foremost position in incidence globally, while ranking second in cancer-related mortality among women [1]. Tailoring treatments for patients with breast cancer entails consideration of various molecular subtypes, with the estrogen receptor (ER) positive subtype constituting the most prevalent [2]. Such patients exhibit high sensitivity to endocrine drugs, enabling endocrine therapy to markedly reduce the risk of breast cancer recurrence and enhance patients' survival rates [3]. Consequently, selective ER modulators have emerged as the cornerstone treatment for ER-positive patients with breast cancer in their early stage [4].

While tissue biopsy remains the primary method for detecting ER expression [5], factors such as the heterogeneity of breast cancer [6], heightened risks of tumor metastasis [7], the need for repeated tumor sampling, time constraints, and relatively high costs can impede the accuracy and representativeness of biopsy results. Hence, there is a pressing need for a simple, effective, and non-invasive approach to assess ER expression status in patients with breast cancer.

Mammography remains the predominant examination method for diagnosing breast cancer. Several studies have indicated a correlation between the features of digital mammography images and ER expression [8-10]. However, the interpretation of these images faces challenges, such as errors from the subjective or experience-dependent judgements of the radiologist, resulting in inconsistent decision-making in clinical treatments between physicians. Machine learning, a technology assisted by computers to process and learn from the input data, holds great capacity in the extraction of high-dimensional features from standard medical images and in the analysis of these features to facilitate clinical decision-making. Preliminary studies have unveiled correlations between quantitative radiomic features extracted from mammography images and molecular biological indices of breast cancer patients [11]. However, conventional radiomics approaches hinge upon precise tumor boundary segmentation, necessitating manual and time-intensive labeling, which can easily overlook crucial radiomic features within the microenvironment and tumor-adjacent tissues. Deep learning (DL)-based radiomics circumvents these challenges with an autonomous learning strategy [12, 13], obviating the necessity for meticulous tumor boundary segmentation processes. Due to their robust feature-learning capabilities, DL models have demonstrated outstanding performance in the detection and classification of breast tumors [14-16]. However, regardless of the substantial advancements made in DL-based classification of breast cancer, existing DL models often struggle to achieve desired performance. For instance, Ueda utilized weighted averages of probability scores from four DL-based models and yielded an area under the receiver operating curves (AUCs) for either ER-positive or -negative breast cancer patients of only 0.67 (0.58-0.76) [17]. Hence, in this study, a novel algorithm was introduced to help build a proficient DL model for strengthening the efficiency of predicting ER expression in breast cancer patients. This algorithm, aiming at extracting ER status-associated information from mammography images for the prediction of ER expression levels in patients with invasive breast cancer, facilitated the construction of the proposed IP-SE-DResNet model.

Materials and methods

Patients

The institutional review board of the First Affiliated Hospital of Nanchang University granted approval to this study. Images were collected from 491 patients diagnosed with invasive breast cancer between October 2016 and October 2021. Digital mammograms were obtained using the Senographe Essential system (GE Medical Systems). Patients were included if: (i) they were diagnosed with primary invasive breast cancer by histological examination, (ii) their ER expression status in tumor samples were confirmed by definitive pathologic examination [18], and (iii) their preoperative mammograms were available. Patients were excluded if: (i) their clinical information was incomplete, (ii) they had undergone preoperative treatment, and (iii) their lesions were entirely blocked by glands or partially showed in mammograms. Ultimately, 358 out of the 491 patients, ranging in age from 22 to 83 years old, with a mean age ± standard deviation of 52.8±6.36, were included in this study. Each patient underwent mammography from both the craniocaudal (CC) view and the mediolateral oblique (MLO) view on the affected side. Tumor samples were acquired through surgical excision. Immunohistochemical testing was conducted, classifying tumor cells were determined as ER-positive if their ER expression level was ≥1% [19], resulting in 227 cases classified as positive and the remaining 131 cases as negative. These ER-positive and -negative designations served as the expected classification labels during the supervised training and testing phases of the IP-SE-DResNet model. Consequently, three datasets encompassing 358 mammography images captured from the CC view (C_Dataset), 358 images from the MLO view (M_Dataset), and 358 images from both

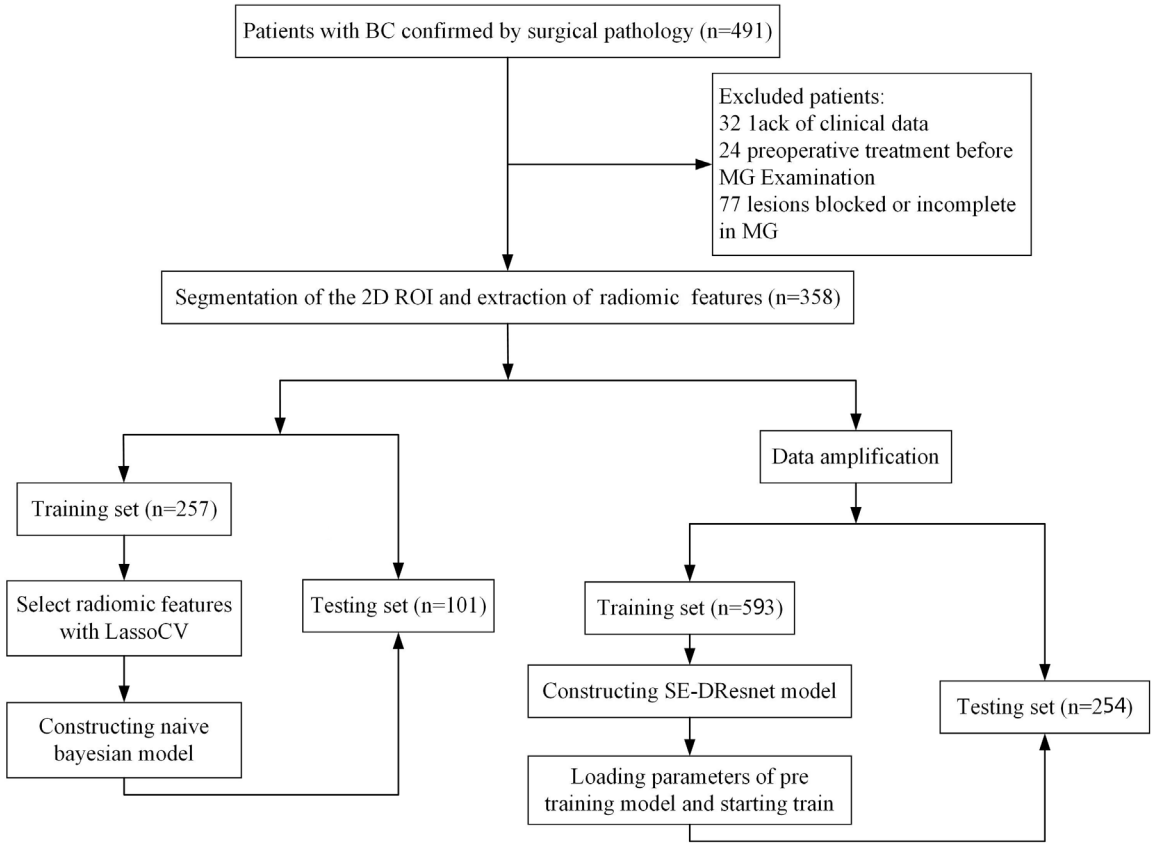

Figure 1. The workflow of the study.

CC and MLO views (CM_Dataset), respectively, were established, with each dataset bearing its own classification label. Within the CM_Dataset, each sample consisted of a two-channel mammogram, where one channel represented the CC view image and the other channel the MLO view image. Subsequently, the three datasets were randomly subdivided into a training dataset (n = 257) and a testing dataset (n = 101), following an approximate 7:3 ratio. Notably, an image sample in the CM_Dataset was generated by combining images of a patient from both the CC and MLO views. The workflow of the study is shown in **Figure 1**. The 7:3 ratio for the division between training and testing datasets is a well-established standard in the literature, commonly employed for this purpose [20].

Data preprocessing

In this study, each dataset contained only 358 images, a sample size that was relatively small. To expand the sample size in each dataset, the mammograph images were rotated or flipped [21]. As a result, a total of 847 images were included in each dataset. The tumors in the images were delineated and labeled by a professional radiologist. The region of the interest (ROI) was determined as the smallest rectangle encompassing the tumor boundary, which was subsequently isolated to undergo data normalization procedures. Following data normalization, the ROI was cropped, with pertinent information including the entire tumor and some surrounding background retained. To facilitate the transfer of initial network parameters from ResNet34 to ResNet50, all ROIs were resized to a resolution of 224×224 pixels. The images were zoomed out using the Cv2.resize function in the Python library. Since the original resolutions of most ROIs were less than 224×224 pixels, the zooming-out process did not result in any information loss. To mitigate information loss caused by the reduction in image resolution, the bilinear interpolation option in the Cv2.resize function was selected for several ROIs sized at 702×931 pixels.

Proposed models

The IP-SE-DResNet model: DL aims to extract inherent abstract representations from data, thereby discerning the mapping relationship between sample data and their corresponding classification labels. Over recent years, notable advancements in DL encompass convolutional neural networks (CNNs) [12, 22], deep residual networks (ResNets) [23, 24], and attention mechanism [25, 26]. CNNs, specialized neural networks, are frequently noted for their computational efficiency and applied in various fields, such as image classification, target detection and semantic segmentation. ResNets are neural networks that employ residual mapping to address complex issues like the "vanishing gradient" in an efficient manner. Common ResNets include ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152 [27-31]. ResNet18 and ResNet34 utilize a basic residual structure, while ResNet50, ResNet101, and ResNet152 are constructed with a bottleneck residual structure. The basic residual structure comprises two 3×3 convolutional layers, maintaining the number of channels in the input feature map unchanged. The bottleneck residual structure comprises a 1×1 convolutional layer, a 3×3 convolutional layer, and a subsequent 1×1 convolutional layer. Here, the 1×1 convolutional layer reduces the number of channels in the feature map, the 3×3 convolutional layer preserves the existing number of channels, and the post 1×1 convolutional layer restores the feature map to a higher number of channels. In CNNs, the incorporation of an attention mechanism mimics human capabilities in identifying salient regions within complex scenes. Existing attention mechanisms encompass various categories, including channel attention, spatial attention, temporal attention, branch attention, channel and spatial attention, and spatial and temporal attention [32]. The SENet [25], which serves as the foundation in our study, pioneered channel attention. At the core of SENet lies the squeeze-and-excitation module, meticulously designed to aggregate global information, emphasize channel-wise relationships, and enhance representational capacity.

In this study, a prediction model that incorporated double ResNets with the squeeze-and-excitation attention mechanism, i.e the IP-SE-DResNet model, was proposed to predict the ER status in patients with breast cancer. The structure of the proposed IP-SE-DResNet model is shown in **Figure 2**.

The proposed IP-SE-DResNet is comprised of the following components and operates as follows:

1. The IP-SE-DResNet combined two sub-networks, SE-ResNet34 and SE-ResNet50, which variants of ResNet34 and ResNet50, respectively, enhanced with the SE attention mechanism, as illustrated in **Figure 2**. The final prediction result was based on the average probability values derived from SE-ResNet34 and SE-ResNet50.

2. The SE-ResNet34 sub-network was comprised of a pre-convolutional layer, a stack of three convolution-attention-residual subnets (CarsNet34), three post-convolutional layers, a fully connected layer, and a softmax. Within this network, the first, second, and third CarsNet34s featured 4, 4, and 6 ResNets, respectively, each containing two convolutional layers equipped with an SE channel attention module. Notably, the final ResNet in each CarsNet34 incorporated a dimension-adjusting shortcut due to a stride of 2 in its initial convolutional layer.

3. The SE-ResNet50 sub-network was comprised of a pre-convolutional layer, a bottleneck-featured ResNet with a channel attention module and a dimension-adjusting shortcut, a series of three convolution-attention-residual subnets (CarsNet50), two post-ResNets featuring a bottleneck design and an SE attention module, a fully connected layer, and a softmax. Within this network, the first, second, and third CarsNet50s included 3, 4, and 6 bottleneck-featured ResNets with an SE attention module, respectively. Notably, the final ResNet in each CarsNet50 incorporated a dimension-adjusting shortcut due to a stride of 2 in its initial convolutional layer.

4. During the network training or testing process, an image sample from the C_dataset, M_dataset, or CM_dataset was fed into both the SE-ResNet34 and SE-ResNet50, which output a probability value suggesting if the image sample was positive or negative. Positive images indicates positive ER status, and vice versa.

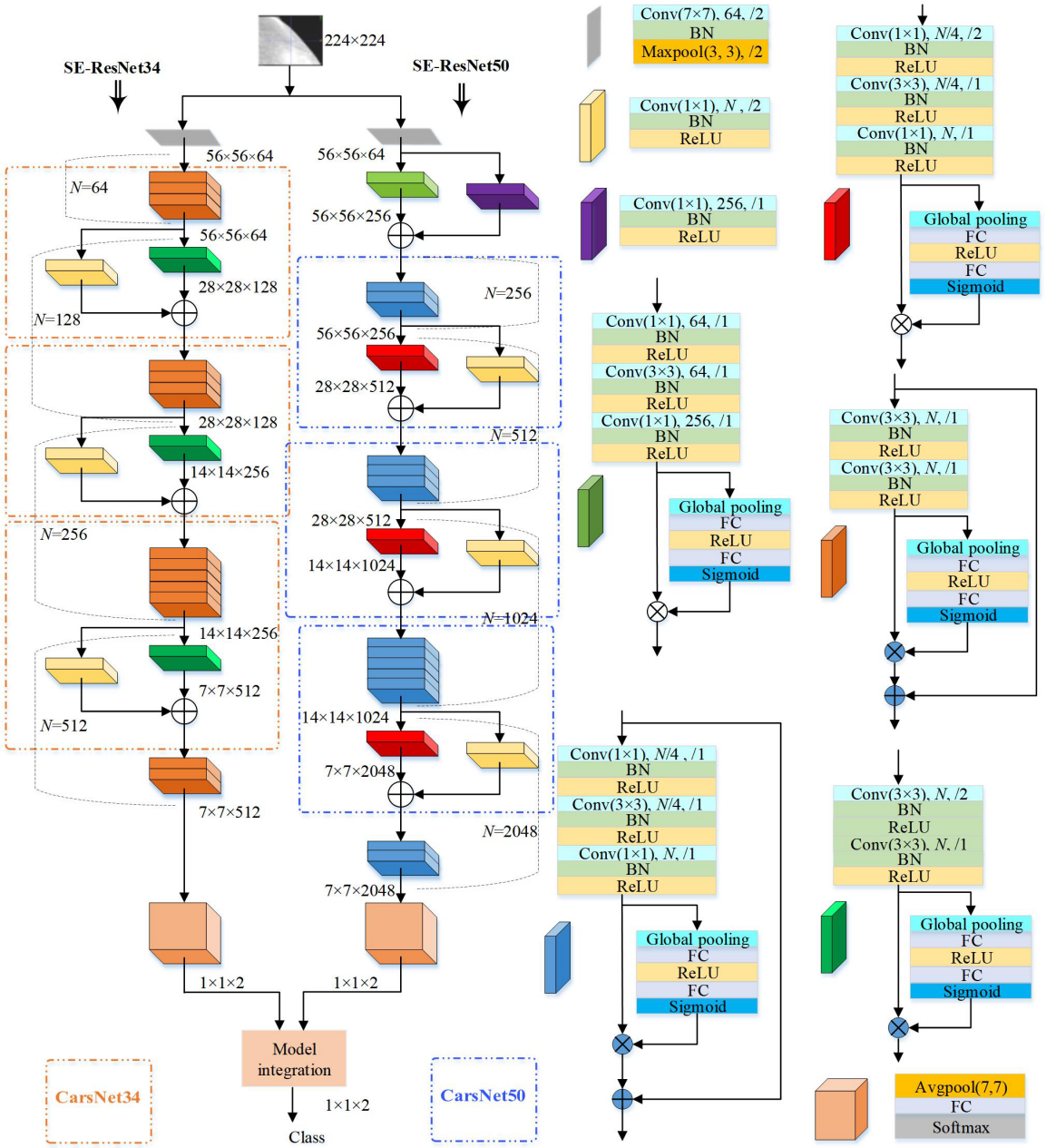

Figure 2. Illustration of the IP-SE-DResNet model. The IP-SE-DResNet model integrated two sub-networks, SE-ResNet34 and SE-ResNet50. During the process of network training or testing, an image sample which contained a minimum rectangular tumor region with a resolution of 224×224 pixels from the C_dataset, M_dataset, or CM_dataset was fed into both the SE-ResNet34 and SE-ResNet50, which output a probability value of the image sample. The average of the two prediction probability values was regarded as the IP-SE-DresNet model's probability output, based on which the final classification label for this image sample was obtained.

The average of two probability values was regarded as the IP-SE-DresNet's probability output, determining the final classification label for the image sample. Generally, a simple average ensemble is sufficient because unless the holdout validation dataset is large and representative, a weighted ensemble is prone to overfitting in comparison to a simple average ensemble.

Radiomic model: The radiomics model [11] is a convectional model employed in the processing and classification tasks for diverse medical images. Early research on radiomics models

generally used radiological characteristics to predict the molecular grading of breast cancer. The following listed the operating procedures of the radiomics models for comparison purpose. First, the preoperative mammography images of patients are randomly divided into a training set and a testing set in a 7:3 ratio. The ROIs was segmented manually by a radiologist, who had been in the practice for over 10 years. Following the segmentation, 95 features were extracted from the breast cancer images using the Pyradiomics toolkit. Subsequently, 22 of the 95 features were selected using the Lasso-CV feature selection method in images captured from the CC, MLO, and CC-and-MLO views. Next, 4 out of the 22 features were screened out by a circular traversal approach, which were fed into the naive Bayesian classifier. Finally, the model that produced the best classification effect was selected, with its features being recorded. The recorded characteristics of the CC view images were CC_Original_glrlm_GrayLevelNonUniformity, CC_original_firstorder_Energy, CC_original_glszm_Sizezone-nonuniformity, and CC_original_gldm_LargeDependenceHighGrayLevelEmphasis. The recorded characteristics of the MLO view images were MLO_original_firstorder_Minimum, MLO_original_shape_MeshVolume, MLO_original_glrlm_GrayLevelNonUniformity, and MLO_original_firstorder_Median. The circular traversal approach was employed again to select 4 features out of the 8 recorded features for exploring features of images from the CC view in conjunction with the MLO view. The characteristics of images from the combined views were CC_original_firstorder_Energy, MLO_original_firstorder_Minimum, MLO_original_shape_MeshVolume, and MLO_original_glrlm_GrayLevelNonUnifor-mity.

Statistical analysis

To begin with, the Kolmogorov-Smirnovtest test was employed to assess the normality of the measurement data. The IBM SPSS software version 17 was utilized for statistical analysis in the study. For measurement data conforming to normal distribution, the independent sample t-test was adopted to assess the significance of the mean age differences between the ER-positive and -negative groups. For counting data, the chi-squared test was used to evaluate the difference in categorical variables, such as breast density and pathological grade, in all cohorts. PyCharm-Python 3.7 and PyTorch 1.9.0 were used as the integrated development environment and deep learning symbolic library, respectively. Positive image sample was defined as ER-positive, and vice versa. According to this definition rule, four result types, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN), were obtained. Next, the prediction accuracy, specificity, and sensitivity of the proposed model were determined according to TP, TN, FP, and FN results. Subsequently, the determined accuracy, specificity, sensitivity, and the area under the receiver operating characteristic curves (AUCs) were used to evaluate the performance of the prediction model. After that, MedCalc was employed to determine the 95% confidence interval of the AUCs. DeLong test was used to compare the AUC values between the proposed model and the radiomics model. *P*-value <0.05 (two-sided) was considered statistically significant.

Results

Patient characteristics

Table 1 outlines the clinical characteristics of the patients. There were no significant differences between the training and testing datasets in terms of age, pathological axillary lymph node (ALN) status, and gland density. However, a statistical difference was observed in the pathological grade between the two datasets, suggesting variability in ER expression levels across different pathological grades in breast cancer patients. This underscores the limitation of local puncture in fully reflecting the pathological basis, which holds significance in clinical treatment.

Prediction performance overview

Table 2 presents the prediction performance of the proposed IP-SE-DResNet model. Performance measurements such as the AUCs, prediction accuracy, sensitivity, and specificity were utilized to evaluate the model's effectiveness. The results suggested that when trained by images from both the CC-and-MLO views, the model performed better in predicting the ER status of breast cancer patients in comparison to that when trained by images from either the CC view or MLO view alone. The AUCs, accu-

Table 1. Clinical characteristics of the 358 breast cancer patients

Clinicopathological features	Training dataset (n = 257)		P	Testing dataset (n = 101)		P
	ER Positive (n = 163)	ER Negative (n = 94)		ER Positive (n = 64)	ER Negative (n = 37)	
Age, mean ± SD, years	52.77±10.57	51.40±9.82	0.306	52.06±9.68	49.43±9.12	0.182
Breast density			0.684			0.717
Fatty	24	10		6	3	
Scattered	57	32		28	14	
Heterogeneous	64	38		22	17	
Extreme	18	14		8	3	
Pathological Grade			0.005			0.019
I	16	3		6	0	
II	120	55		47	23	
III	27	36		11	14	
Pathological ALN status			0.744			0.573
Positive	85	51		39	21	
Negative	78	43		25	16	

ALN, pathological axillary lymph node; ER, Estrogen receptor.

Table 2. Comparison of performance measurements between the IP-SE-DResnet model and the radiomics model

Method	Training dataset (n = 593)				Testing dataset (n = 254)			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Radiomics (CC)	0.590	0.684	0.530	0.607	0.552	0.833	0.365	0.599
Radiomics (MLO)	0.552	0.540	0.573	0.576	0.590	0.392	0.704	0.570
Radiomics, CC+MLO	0.635	0.430	0.750	0.614	0.571	0.314	0.815	0.613
IP-SE-DResNet, CC	0.816	0.765	0.872	0.849	0.735	0.777	0.690	0.835
IP-SE-DResNet, MLO	0.836	0.790	0.884	0.858	0.767	0.744	0.791	0.846
IP-SE-DResNet CC-and-MLO	0.882	0.815	0.954	0.895	0.831	0.810	0.850	0.886

CC, cranio caudal; MLO, mediolateral oblique.

racy, sensitivity, and specificity values were 0.895 (95% CIs: 0.886-0.913), 0.882, 0.815, and 0.954 in the training dataset and 0.886 (95% CIs: 0.809-0.934), 0.831, 0.810, and 0.850 in the testing dataset, respectively.

Comparison between the different datasets from different views

This research was based on transfer learning. For comparison, the proposed model was trained separately with three datasets, the C_ Dataset, the M_dataset, and the CM_Dataset. Specifically, the proposed model underwent training from scratch each time by being transitioned to a pre-trained status (ResNet34 and NesNet50). This transfer guaranteed a fresh start for every training process, preventing any potential interference between different training processes. **Figure 3** shows the ROCs for the training and testing datasets. According to results presented in **Figure 3**, all ROCs yielded by the proposed model when trained by images from the CC, MLO, and CC-and-MLO views were relatively smooth, indicating that the IP-SE-DResNet model was unlikely to be over-fitted. The AUCs of the model when trained by images from the CC, MLO, and CC-and-MLO views in the training and testing datasets were 0.849 (95% CIs: 0.809-0.868) and 0.835 (95% CIs: 0.790-0.887) (CC views alone), 0.858 (95% CIs: 0.813-0.872) and 0.846 (95% CIs: 0.793-0.889) (MLO views alone), and 0.895 (95% CIs: 0.886-0.913) and 0.886 (95% CIs: 0.809-0.934) (CC-and-MLO views), respectively. It was observed that the AUCs for the proposed model

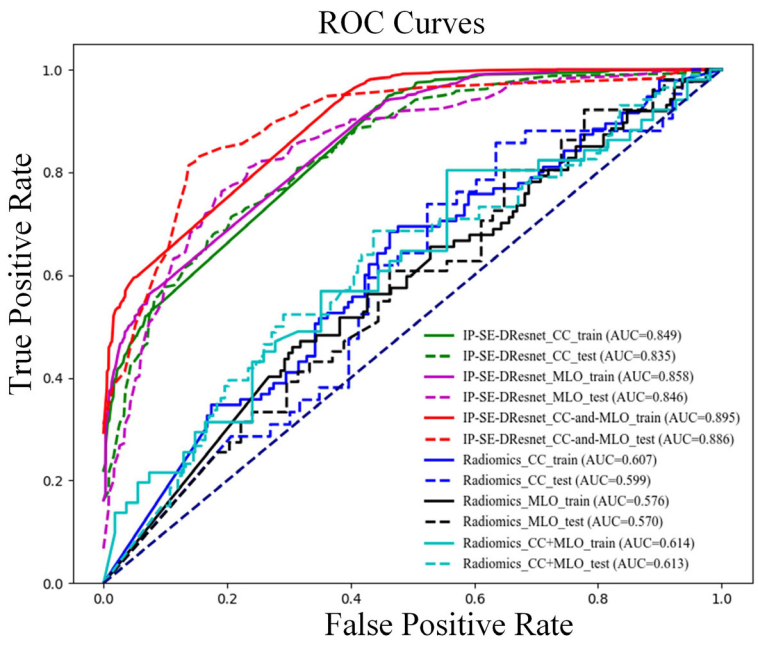

Figure 3. Comparison of the ROCs between the IP-SE-DResnet model and the radiomics model.

when trained by images from the CC-and-MLO views in both the training and testing datasets were extremely close and the highest, indicating that the application of images from the CC-and-MLO views for model training achieved more accurate results in predicting the ER status in breast cancer patients than the employment of images captured from either the CC view or the MLO view alone. The DeLong test results showed no significant differences in the comparison of AUCs between the C_Dataset and the M_dataset (P = 0.075). However, statistically significant differences were observed in the comparison of AUCs between the C_Dataset and the CM_dataset, as well as between the M_Dataset and the CM_dataset (P<0.0001).

Comparison of prediction performance between the IP-SE-DResNet model and the radiomics model

The quantitative performance comparison outlined in **Table 2** indicates that the proposed IP-SE-DResNet model exhibited superior performance in predicting ER status in breast cancer patients compared to the radiomics model based on the naive Bayesian classifier across all three sub-datasets that encompassed images from the CC view, the MLO view, and the CC-and-MLO view. The IP-SE-DResNet model demonstrated higher accuracy in predicting ER status in both the training and testing datasets than the traditional radiomics model did. In addition, the proposed model consistently outperformed the radiomics model in all evaluation measurements, with the AUC for the former exceeding that for the latter by as high as 27 percent (P<0.001).

Suspicious tumor area discovery

Figure 4 depicts the suspicious areas identified by the proposed IP-SE-DResNet model. When the mammography images of breast cancer patients were fed into the IP-SE-DresNet model, an attention map was thus generated. This attention map was extracted from the last layer of the first convolution-attention-residual subnets (CarsNet34) in the SE-ResNet34 sub-network. It showed the attention degree of the areas in the input images that attracted the attention of the model. These areas were important because they were in direct association with the ER status in breast cancer patients. Therefore, these areas were considered suspicious that required further investigation. In the study, 0.5 was the threshold value for pixels in the attention map to preserve the suspicious areas. It was a standard derived from previous report [33].

Ablation studies

Ablation studies were performed to analyze the distinct contributions of various sub-networks in the selection process. The pre-trained ResNet18, ResNet34, ResNet5, ResNet101, and ResNet152 models were downloaded from the PyTorch official website. According to the principle of transfer learning, the pre-trained network parameters of all the above ResNets models were loaded into the IP-SE-DResNet model to replace the previous

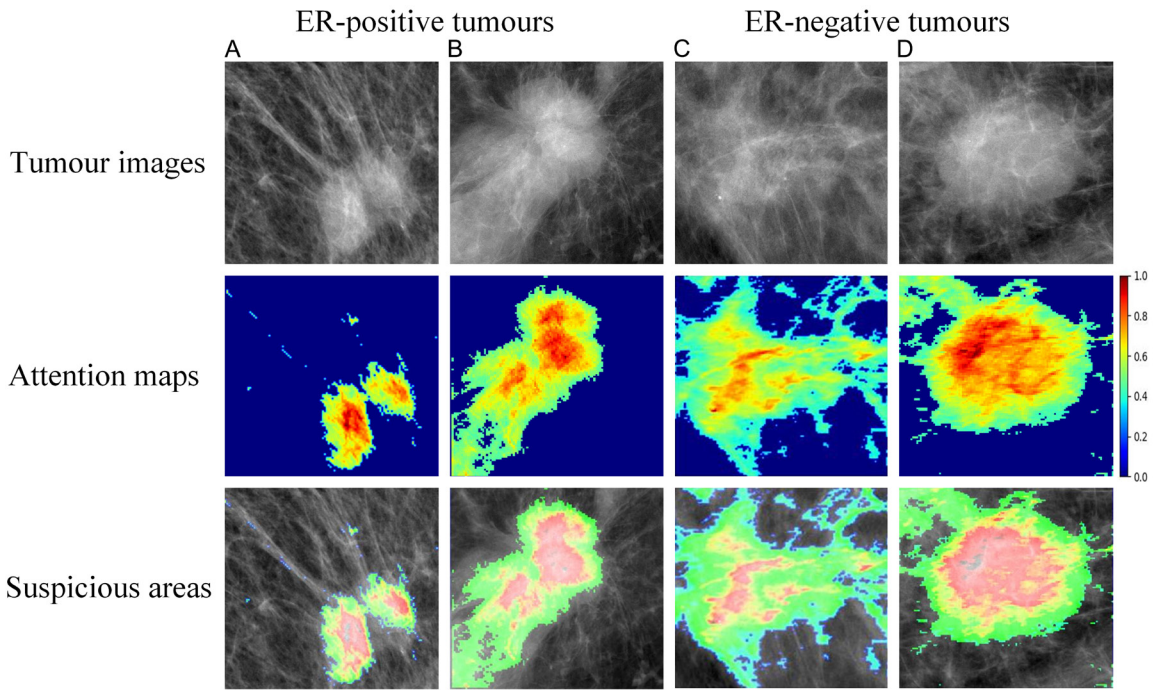

Figure 4. Suspicious tumour area discovery. The attention maps show the attention degree of those areas in the input tumor images that attract the attention of the proposed model. These suspicious areas are directly related to the ER status. ER, Estrogen receptor.

Table 3. Comparison of prediction performances between different SE-Resnets

Method	View	Accuracy	Precision	Recall	F1-score	AUC
SE-Resnet18	CC	0.714	0.705	0.769	0.735	0.771
	MLO	0.682	0.680	0.719	0.699	0.774
	CC-and-MLO	0.703	0.657	0.828	0.733	0.774
SE-Resnet34	CC	0.731	0.742	0.736	0.739	0.788
	MLO	0.708	0.728	0.686	0.706	0.785
	CC-and-MLO	0.746	0.780	0.672	0.722	0.806
SE-Resnet50	CC	0.726	0.748	0.711	0.729	0.785
	MLO	0.703	0.714	0.702	0.708	0.783
	CC-and-MLO	0.746	0.818	0.621	0.706	0.793
SE-Resnet101	CC	0.731	0.723	0.777	0.749	0.787
	MLO	0.699	0.712	0.694	0.703	0.779
	CC-and-MLO	0.746	0.726	0.776	0.750	0.790
SE-Resnet152	CC	0.705	0.802	0.570	0.667	0.782
	MLO	0.699	0.727	0.661	0.693	0.782
	CC-and-MLO	0.695	0.667	0.759	0.719	0.784

CC, cranio caudal; MLO, mediolateral oblique.

ResNet34 or/and SE-ResNet50 parameters. The performances of the model with different sub-network configurations are presented in **Table 3**. From **Table 3**, we selected the top two networks with the highest AUC to be included in the modeling of this study.

Discussion

Over recent years, there has been a growing body of research examining the correlation between ER status and mammogram features. Brigid K. et al. found no statistical difference

between ER status and the most common X-ray features, such as masses and calcifications. However, they did observe a positive correlation between ER expression and the presence or absence of axillary lymph node enlargement and nipple depression [34]. Other studies showed that ER-negative breast cancer patients had a higher risk of mammographic density in comparison to ER-positive breast cancer patients [35, 36]. In our study, there is no statistical difference in mammographic density and pathological axillary lymph node between ER-negative and -positive patients. Unlike the findings of Tailaiti, et al.'s [34], it is possible that this is due to racial or geographic differences. Yan Li et al. reported that the ER-positive cancer patients showed higher propensity for association with mammography image classification in comparison to HER2-positive cancer patients [37]. Another study indicated that the ER positiveness was associated with architectural distortion [38]. However, despite these findings indicating a specific correlation between ER status and mammogram features, the results were notably limited in sample size and lacked accuracy [39].

Radiomics methods analyze tumor information at the macroscopic level by mining quantitative image features that are associated with ER status from mammograph images. While the radiomics methods have their advantages, as the extracted features were considered distinctive, their reliance on expert-defined features means they may not represent the optimal feature quantification method for the imminent differentiated tasks [40]. Meanwhile, the robustness of the prediction process affects the effectiveness of the training process and the reliability of the prediction results. In our study, compared with the radiomics model that only obtained an AUC/accuracy of 0.613/0.571 in the testing datasets for determining ER status, the proposed IP-SE-DResNet model acquired only user-defined and easy-to-crop rectangular ROIs, accompanied by reduction in dependence on manual selection of the optimizer and classifier. ER-related tumor features were automatically acquired from the large-scale datasets with the use of DL approaches. In the proposed IP-SE-DResNet model, the determination of ER status achieved high accuracy, with the model trained by images from the CC-and-MLO views showing optimal performance with an AUC/accuracy of 0.886/0.831 in the testing dataset ($P<0.001$). This finding is consistent with the results of Zhou et al. [41]. In their study on the prediction of HER2 expression using imaging histology, in which a model combining the imaging histologic features of 2 photographic positions predicted ER expression better than a model with single position imaging histologic features.

Unlike Ueda's hybrid DL model [17], which integrated four DL models, our model opted for simplicity by selecting merely two ResNets models with straightforward structures for integration. This decision was made to avoid the heavy computational burden and overfitting associated with a large number of network parameters and excessive calculations, which could potentially compromise classification performance. Furthermore, we proposed the application of an attention mechanism, specifically the SE modile, to the channel level of the ResNet. This approach allowed for the assignment of varying weights to different channels of the feature map. The SE module's non-linearity enabled an enhanced fitting of the complex correlation between channels, resulting in improved performance. In our study, the proposed IP-SE-DResNet model fed with mammograms to predict ER expression status in patients with primary invasive breast cancer showed promising results. Our study revealed a significant association between mammogram features and ER status in breast cancer patients. In addition, the IP-SE-DResNet model offers the advantage of requiring unprocessed tumor mammograms. Its non-invasive nature, coupled with its straightforward operation, allows for direct prediction of ER status. The model's results serve as a supplementary tool to biopsy examination findings for patients undergoing their first operation. Additionally, the IP-SE-DResNet model can predict the tumor region most strongly associated with ER status. This analysis helps clinicians intuitively understand the predicted results in mammograms.

Despite the encouraging results from the IP-SE-DResNet model, the study still has some limitations. Firstly, it's worth noting that all data in our study were collected from a single center. To ensure the reproducibility and generalizability of the model, future research

should incorporate data from multiple sources or consider a prospective study design. Secondly, our study focused only on ER status. In future research, it is imperative to explore the relationship between the radiomics features of breast cancer and other prognostic factors such as PR, Her-2, and Ki-67. Lastly, given the small dataset size in our study, larger and high-quality datasets are necessary to enhance the robustness of findings in future investigations. In addition, an automatic ROI segmentation is expected in future work.

Acknowledgements

This study was supported by the Education Department Project of Jiangxi Province, China (Grant No. GJJ2200132) and Clinical Research Center for Medical Imaging in Jiangxi Province (No. 20223BCG74001).

Disclosure of conflict of interest

None.

Address correspondence to: Dechang Peng, Department of Radiology, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China. Tel: +86-0791-83969675; E-mail: pengdcdoctor@163.com; Jianhua Wu, School of Information Engineering, Nanchang University, Nanchang, Jiangxi, China. Tel: +86-0791-88603000; E-mail: jhwu@ncu.edu.cn

References

[1] Siegel RL, Miller KD, Fuchs HE and Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021; 71: 7-33.

[2] Chen P, Li B and Ou-Yang L. Role of estrogen receptors in health and disease. Front Endocrinol (Lausanne) 2022; 13: 839005.

[3] Kurebayashi J. Current clinical trials of endocrine therapy for breast cancer. Breast Cancer 2007; 14: 200-214.

[4] Waks AG and Winer EP. Breast cancer treatment: a review. JAMA 2019; 321: 288-300.

[5] Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, Hayes DF, Lakhani SR, Chavez-MacGregor M, Perlmutter J, Perou CM, Regan MM, Rimm DL, Symmans WF, Torlakovic EE, Varella L, Viale G, Weisberg TF, McShane LM and Wolff AC. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. J Clin Oncol 2020; 38: 1346-1366.

[6] Januškevičienė I and Petrikaitė V. Heterogeneity of breast cancer: the importance of interaction between different tumor cell populations. Life Sci 2019; 239: 117009.

[7] Santiago L, Adrada BE, Huang ML, Wei W and Candelaria RP. Breast cancer neoplastic seeding in the setting of image-guided needle biopsies of the breast. Breast Cancer Res Treat 2017; 166: 29-39.

[8] Shaikh AJ, Mullooly M, Sayed S, Ndumia R, Abayo I, Orwa J, Wasike R, Moloo Z and Gierach GL. Mammographic breast density and breast cancer molecular subtypes: the Kenyan-African aspect. Biomed Res Int 2018; 2018: 6026315.

[9] Sturesdotter L, Sandsveden M, Johnson K, Larsson AM, Zackrisson S and Sartor H. Mammographic tumour appearance is related to clinicopathological factors and surrogate molecular breast cancer subtype. Sci Rep 2020; 10: 20814.

[10] Adrada BE, Huo L, Lane DL, Arribas EM, Resetkova E and Yang W. Histopathologic correlation of residual mammographic microcalcifications after neoadjuvant chemotherapy for locally advanced breast cancer. Ann Surg Oncol 2015; 22: 1111-1117.

[11] Ma W, Zhao Y, Ji Y, Guo X, Jian X, Liu P and Wu S. Breast cancer molecular subtype prediction by mammographic radiomic features. Acad Radiol 2019; 26: 196-201.

[12] Lecun Y, Bengio Y and Hinton G. Deep learning. Nature 2015; 521: 436-444.

[13] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T and Hassabis D. Mastering the game of Go without human knowledge. Nature 2017; 550: 354-359.

[14] Abdel-Zaher AM and Eldeib AM. Breast cancer classification using deep belief networks. Exp Systems with Applications 2016; 46: 139-144.

[15] Ha R, Chang P, Mema E, Mutasa S, Karcich J, Wynn RT, Liu MZ and Jambawalikar S. Fully automated convolutional neural network method for quantification of breast MRI fibroglandular tissue and background parenchymal enhancement. J Digit Imaging 2019; 32: 141-147.

[16] Zhou J, Luo LY, Dou Q, Chen H, Chen C, Li GJ, Jiang ZF and Heng PA. Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J Magn Reson Imaging 2019; 50: 1144-1151.

[17] Ueda D, Yamamoto A, Takashima T, Onoda N, Noda S, Kashiwagi S, Morisaki T, Honjo T, Shimazaki A and Miki Y. Training, validation, and test of deep learning models for classification of receptor expressions in breast cancers from mammograms. JCO Precis Oncol 2021; 5: 543-551.

[18] Allred DC, Carlson RW, Berry DA, Burstein HJ, Edge SB, Goldstein LJ, Gown A, Hammond ME, Iglehart JD, Moench S, Pierce LJ, Ravdin P, Schnitt SJ and Wolff AC. NCCN task force report: estrogen receptor and progesterone receptor testing in breast cancer by immunohistochemistry. J Natl Compr Canc Netw 2009; 7 Suppl 6: S1-S21; quiz S22-23.

[19] Fusco N, Ragazzi M, Sajjadi E, Venetis K, Piciotti R, Morganti S, Santandrea G, Fanelli GN, Despini L, Invernizzi M, Cerbelli B, Scatena C and Criscitiello C. Assessment of estrogen receptor low positive status in breast cancer: implications for pathologists and oncologists. Histol Histopathol 2021; 36: 1235-1245.

[20] Qi Q, Li Y, Wang J, Zheng H, Huang Y, Ding X and Rohde GK. Label-efficient breast cancer histopathological image classification. IEEE J Biomed Health Inform 2019; 23: 2108-2116.

[21] Bloice MD, Roth PM and Holzinger A. Biomedical image augmentation using augmentor. Bioinformatics 2019; 35: 4522-4524.

[22] Krizhevsky A, Sutskever I and Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM 2017; 60: 84-90.

[23] Wang H, Wang J, Xu H, Sun Y and Yu Z. DRSN-Fuse: deep residual Shrinkage Network for infrared and visible image fusion. Sensors 2022; 22: 5149.

[24] He K, Zhang X, Ren S and Sun J. Identity mappings in deep residual networks//Computer Vision-ECCV 2016. 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV 14. Springer International Publishing; 2016. pp. 630-645.

[25] Hu J, Shen L, Albanie S, Sun G and Wu E. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 2020; 42: 2011-2023.

[26] Papanastasiou G, Dikaios N, Huang J, Wang C and Yang G. Is attention all you need in medical image analysis? A review. IEEE J Biomed Health Inform 2024; 28: 1398-1411.

[27] Yu X and Wang SH. Abnormality diagnosis in mammograms by transfer learning based on ResNet18. Fundamenta Informaticae 2019; 168: 219-230.

[28] Korfiatis P, Kline TL, Lachance DH, Parney IF, Buckner JC and Erickson BJ. Residual deep convolutional neural network predicts MGMT methylation status. J Digit Imaging 2017; 30: 622-628.

[29] Yang Q, Jiang S, Chen J and Lin W. Crack detection based on ResNet with spatial attention. Computers and Concrete, An International Journal 2020; 26: 411-420.

[30] Ahuja S, Panigrahi BK, Dey N, Rajinikanth V and Gandhi TK. Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. Appl Intell (Dordr) 2021; 51: 571-585.

[31] Singh D, Kumar V and Kaur M. Densely connected convolutional networks-based COVID-19 screening model. Appl Intell (Dordr) 2021; 51: 3044-3051.

[32] Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ, Zhang SH, Martin RR, Cheng MM and Hu SM. Attention mechanisms in computer vision: a survey. Computational Visual Media 2022; 8: 331-368.

[33] Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, Liu Y, Gevaert O, Wang K, Zhu Y, Zhou H, Liu Z and Tian J. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. Eur Respir J 2019; 53: 1800986.

[34] Tailaiti G, Maimaiti G, Aikeremu Y and Tuerdi B. Molybdenum target X-ray features and estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 in invasive breast cancer. Int J Gen Med 2021; 14: 2777-2783.

[35] Razzaghi H, Troester MA, Gierach GL, Olshan AF, Yankaskas BC and Millikan RC. Association between mammographic density and basal-like and luminal A breast cancer subtypes. Breast Cancer Res 2013; 15: R76.

[36] Shieh Y, Scott CG, Jensen MR, Norman AD, Bertrand KA, Pankratz VS, Brandt KR, Visscher DW, Shepherd JA, Tamimi RM, Vachon CM and Kerlikowske K. Body mass index, mammographic density, and breast cancer risk by estrogen receptor subtype. Breast Cancer Res 2019; 21: 48.

[37] Li Y, Cao J, Zhou Y, Mao F, Shen S and Sun Q. Mammographic casting-type calcification is an independent prognostic factor in invasive breast cancer. Sci Rep 2019; 9: 10544.

[38] Killelea BK, Chagpar AB, Bishop J, Horowitz NR, Christy C, Tsangaris T, Raghu M and Lannin DR. Is there a correlation between breast cancer molecular subtype using receptors as surrogates and mammographic appearance? Ann Surg Oncol 2013; 20: 3247-3253.

[39] Conti A, Duggento A, Indovina I, Guerrisi M and Toschi N. Radiomics in breast cancer classification and prediction. Semin Cancer Biol 2021; 72: 238-250.

[40] Hosny A, Parmar C, Quackenbush J, Schwartz LH and Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018; 18: 500-510.

[41] Zhou J, Tan H, Bai Y, Li J, Lu Q, Chen R, Zhang M, Feng Q and Wang M. Evaluating the HER-2 status of breast cancer using mammography radiomics features. Eur J Radiol 2019; 121: 108718.