Biomedical Signal Processing and Control 73 (2022) 103453

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Discrimination of benign and malignant solid breast masses using deep residual learning-based bimodal computer-aided diagnosis system

Zahra Assari [a], Ali Mahloojifar [a,*], Nasrin Ahmadinejad [b]

[a] Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
[b] Medical Imaging Center, Cancer Research Institute, Imam Khomeini Hospital Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Tehran University of Medical Sciences (TUMS), Tehran, Iran

ARTICLE INFO

Keywords:
Solid breast mass
Mammography
Ultrasound imaging
Bimodal computer-aided diagnosis system
Deep learning
Residual learning

ABSTRACT

One of the most common breast cancer mammographic manifestation is solid mass. If the information obtained from mammography is inadequate, complementary modalities such as ultrasound imaging are used to achieve additional information. Although interest in the combination of information from different modalities is increasing, it is an extremely challenging task. In this regard, a computer-aided diagnosis (CAD) system can be an efficient solution to overcome these difficulties. However, most of the studies have focused on the development of mono-modal CAD systems, and a few existing bimodal ones rely on the extracted hand-crafted features of mammograms and sonograms. In order to meet these challenges, this paper proposes a novel bimodal deep residual learning model. It consists of the following major steps. First, the informative representation for each input image is separately constructed. Second, in order to construct the high-level joint representation of every two input images and effectively explore complementary information among them, the representation layers of them are fused. Third, all of these joint representations are fused to obtain the final common representation of the input images for the mass. Finally, the recognition result is obtained based on information extracted from all input images. The augmentation strategy was applied to enlarge the collected dataset for this study. Best recognition results on the sensitivity, specificity, F1-score, area under ROC curve, and accuracy metrics of 0.898, 0.938, 0.916, 0.964, and 0.917, respectively, are achieved by our model. Extensive experiments demonstrate the effectiveness and superiority of the proposed model over other state-of-the-art models.

1. Introduction

The most common breast cancer mammographic manifestations are mass and microcalcification. In breast image analysis, it is a very challenging task to distinguish malignant masses from benign ones. This is due to some issues such as the high degree of visual similarity between them, the relatively low contrasts and obscure boundaries between masses and normal surrounding tissue, the large variability of masses. A breast biopsy is the diagnostic procedure that can definitely determine whether a suspicious solid mass is benign or malignant. However, this method has also numerous complications [1–6]. Therefore, in order to reduce the number of unnecessary biopsies, the use of non-invasive imaging techniques is necessary. Currently, among all these techniques, mammography is considered as the gold standard for the early detection and diagnosis of breast cancer. Nevertheless, it suffers from its limitations. If the information obtained from mammography is inadequate, complementary modalities such as ultrasound (US) imaging are suggested to the patient to achieve additional information. US imaging, as a valuable complement to mammography for further evaluation of suspicious solid masses, is widely available, requires no contrast agent, well tolerated, relatively inexpensive [7].

Radiologists have found that the combination of information from multiple modalities, can lead to more effective diagnosis [8]. However, it is an extremely challenging task. The Breast Imaging Reporting and Data System (BI-RADS) lexicon [9] is widely used by the radiologists to characterize the masses seen in the sonograms and mammograms. This lexicon includes five sonographic and three mammographic descriptors for the breast masses. The radiologists classify them into one of the BI-RADS assessment categories based on their described features. Each category has a specific required action. For instance, if a BI-RADS 3 mass shows no change in the short-interval follow-up study, the final assessment is changed to BI-RADS 2 (benign) and there's no need to do a

* Corresponding author.
E-mail addresses: z.assari@modares.ac.ir (Z. Assari), mahlooji@modares.ac.ir (A. Mahloojifar).

https://doi.org/10.1016/j.bspc.2021.103453
Received 17 March 2021; Received in revised form 23 November 2021; Accepted 16 December 2021
Available online 27 December 2021

biopsy. In fact, the radiologists mentally integrate the complementary information from the sonograms and mammograms acquired from a patient for further evaluation of the suspicious mass. Nevertheless, this mental integration of information is usually time-consuming, error-prone and subjective. All the factors previously mentioned, as shown in Fig. 1, make the fusion of information from various modalities a challenging task.

A CAD system can be developed to help radiologists in facilitating data fusion from different modalities and play a critical role in improving the diagnostic performance of breast cancer [10]. However, most of the studies have focused upon the development of mono-modal CAD systems. Therefore, a lot of efforts still need to be dedicated to designing bimodal and multimodal ones. In spite of the increasing need for these types of systems, few studies have been done to distinguish malignant breast lesions from benign ones using the combination of information from different modalities [7], [2],[11], [12].

Some researchers attempted to apply hand-crafted features to classify the lesions. Drukker et al. [2] developed a CAD system to distinguish malignant lesions from benign ones. They performed the analysis on their own dataset of 100 lesions (40 malignant, 40 benign solid, and 20 cystic lesions). The initial features included five features for mammograms (pertaining to shape, density, texture, margin sharpness, and spiculation) and four features for US images (pertaining to shape, texture, margin, and posterior acoustic behavior). There was an average of three mammograms and two sonograms per lesion. For each feature, its mean, maximum, and minimum values were considered for all images representing a lesion. Thus, the total number of features extracted from a lesion was equal to 27. They utilized the combination extracted features of lesions from mammograms and sonograms, and the Bayesian neural network. A subset of features was selected based on automated stepwise optimization. Results showed a statistically significant increase over that achieved with features from either mammography or ultrasound alone. Sidiropoulos et al. [11] designed a GPU-based CAD system to accurately discriminate benign to malignant lesions. They used their own dataset of 62 histologically verified lesions (32 malignant and 30 benign cases). The lesions were outlined on the images by an experienced radiologist. Each lesion was represented by an eighty-feature vector, which consisted of 40 textural features extracted from each image (i.e., the mammogram and sonogram). These features comprised four features from the ROI's histogram, twenty-six features from the ROI's co-occurrence matrices, and 10 features from the ROI's run-length matrices. All features were normalized to zero mean and unit standard deviation. They utilized the parallel programming, the probabilistic neural network classifier, and the extracted textural features from mammograms and sonograms. However, these features contain limited information and lack high-level representations [13–15]. Therefore, they cannot adapt well for the complex solid mass recognition [7].

In the last decade, convolutional neural networks (CNNs) have been demonstrated to yield highly discriminative representations that have aided in many medical image analysis tasks [16–18,7,19,20,14,15,21]. They can exploit the higher-level features from the lower-level ones hierarchically [13–15], and thus achieving superior performances compared with the hand-crafted feature extraction methods [7].

Carneiro et al. [22] developed a CAD system for multi-view classification. A separate CNN was trained for each view (i.e., CC and MLO views), initially. Then, using the features learned from them, a final CNN was trained that estimates the patient's risk of developing cancer. Arevalo et al. [23] presented a representation learning CNN-based framework for mammography mass classification. Cheng et al. [24] performed a comprehensive study on the deep-learning-based CAD systems for the differential diagnosis of benign and malignant breast lesions. Huynh et al. [25] demonstrated the potential usefulness of transfer learning for mammographic tumor classification. Lévy and Jain [26] presented how a shallow CNN, an AlexNet, and a GoogleNet can be used to classify pre-segmented masses in mammograms, directly, using a combination of transfer learning, careful pre-processing and data augmentation. Wang et al. [27] utilized the stacked autoencoder-based model for discrimination of masses with microcalcifications on mammography. Han et al. [28] modified the GoogleNet model to differentiate benign and malignant lesions in US images. Sun et al. [29] developed a graph based semi-supervised learning scheme using CNN for mammography mass

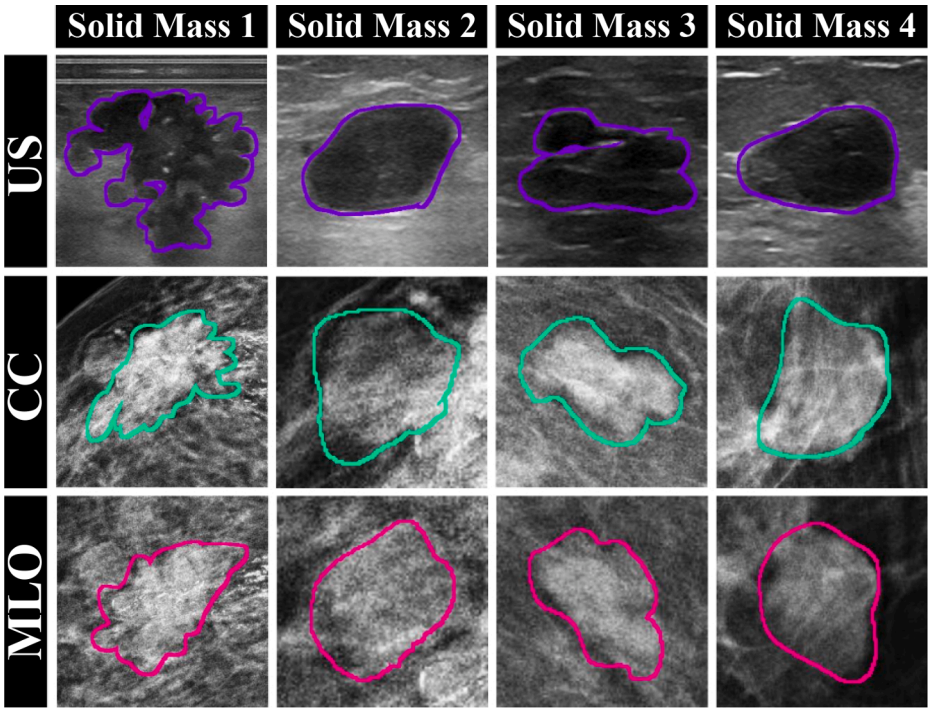

Fig. 1. Representative sonographic and both standard mammographic images of four solid breast masses. Each column represents one mass. Examples in the first two columns are malignant, while ones in the last two columns are benign. Blue, green, and red contours indicate the solid masses in the ultrasound (US), CranioCaudal (CC), and MedioLateral Oblique (MLO) images, respectively.

classification. Shen et al. [30] compared the VGG and ResNet models to detect cancer on screening mammograms. Yu et al. [31] proposed a semi-automatic CAD system using DenseNet201 to classify mammograms into the normality and abnormality classes. The analysis was performed on the public MINI-MIAS database. They used two transfer learning methods and explored which method would work better in their work. The comparison results between their model and the other five state-of-the-art networks show that their method achieved the highest accuracy. Moon et al. [32] employed the ensemble method to combine multiple CNN results. The VGG, ResNet, and DenseNet models were utilized. In their study, the best recognition results were achieved by the DenseNet-based model to classify masses into benignancy or malignancy on breast US images. Zhang et al. [33] developed a new model called BDR-CNN-GCN to improve detection of malignant lesions in mammograms. This model is a combination of a two-layer graph convolutional network (GCN) and a BDR-CNN model, which is a combination of a standard 8-layer CNN, batch normalization (BN), drop out (DO), and rank-based stochastic pooling (RSP). The GCN aids to extract relation-awareness features, whereas the BDR-CNN aids to extract image-level features. The experiments showed the BDR-CNN-GCN model was superior to 15 state-of-the-art methods.

Images of different modalities collected from the same mass could provide much richer recognition information than from a single one. Habib et al. [34] developed a bimodal GoogleNet-based CAD system to distinguish malignant breast lesions from benign ones. First, a distinct pre-trained GoogleNet model was fine-tuned separately for each modality (i.e., the mammography and ultrasound), generating high-level feature vectors. Then, a concatenation of these feature vectors originating from mammograms and sonograms was used to train a fully connected network architecture to provide the final classification. They used their own dataset of 153 biopsy-proven lesions, consisting of 73 malignant and 80 benign cases. The experimental results showed improvement in AUC when utilizing both modalities (i.e., the mammography and ultrasound imaging). However, most of the existing CAD systems are designed for images of a single modality in breast mass classification [7,1]. Therefore, applying CNNs to solid mass classification still needs specifically designed network architecture to extract complex correlations of multimodal images [34].

How to implement a combination of the images' features of different modalities properly for more accurate classification is becoming a critical task. The existing fusion strategies can be categorized into three forms in general: (1) input fusion, (2) result fusion, and (3) feature-maps fusion. The input fusion strategy simply serializes input images together as a 4-D input feature vector or combines them with fixed weights, regardless of the characteristics of different images. Actually, this crude strategy even cannot guarantee the improvement after the fusion task. The result fusion strategy predicts the classification results using each image independently and then combines their results. Although this strategy can partly achieve superior performance compared with the input fusion strategy, it fails to extract the complex intrinsic correlations between input images. Feature-maps fusion is a robust and efficient strategy. The aim is to combine feature maps extracted from the input images into a single feature map having more discriminant information than the input feature maps [13].

The main contributions of our work are summarized as follows:

- Due to exploit the high-level features, deep learning-based models can achieve significantly superior performances compared with the hand-crafted feature based CAD systems. Residual learning techniques are utilized to cope with the degradation and overfitting problems of deeper networks. Therefore, they allow the models to benefit the more depth for more accurate recognition. Thus, in this paper, a novel bimodal deep residual learning model is presented to meet the challenges of the information combination from the images of two modalities (the mammography and US imaging) for solid breast mass classification.
- This study has focused on the feature-maps fusion to implement suitable combinations of the extracted deep features of input images. In order to reduce the computational costs and improve the generalization of the model, a transfer learning strategy has been exploited. For the proposed bimodal framework, six different configurations were considered and evaluated to find the configuration with better performance. The results demonstrated that the higher-level feature maps are more suitable for feature-maps fusion.
- This study focuses on a specific type of mass (i.e., solid mass). Inspired by the approach of radiologists, the two standard mammographic views and sonographic images are considered as separate inputs. The proposed model is trained for all input images simultaneously. The database used in this study comprised the images from female patients with a total of 156 solid breast masses, of which 77 were malignant and 79 were benign. Artificial data augmentation is a solution widely used in the context of CNN, for increasing the number of datasets and reduce overfitting. Therefore, three augmentation operators were applied to enlarge the training dataset.
- It is necessary to investigate whether the proposed CAD system can enhance the solid mass classification in comparison with representative models in the literature. Therefore, it has been compared with several hand-crafted features based bimodal CAD systems and recently published deep learning-based models. These experiments demonstrate the effectiveness and superiority of the proposed framework compared with other state-of-the-art models. To the best of our knowledge, the proposed model is among the earliest efforts of applying deep learning approaches to design bimodal CAD system. This may inspire more studies to solve challenging multi-modal medical image analysis problems.

2. Materials and methods

2.1. Database

Institutional review board approval was obtained for this study. According to the BI-RADS lexicon [9], women with solid masses who had biopsy-proven disease or were determined to involve benign masses through short-interval follow-up study were eligible. Between August 2018 and October 2020, the acquired mammograms and sonograms were collected using a FUJIFILM Digital Mammography System (FDR-3000AWS, FUJIFILM Corporation, 26–30, Nishiazabu 2-chome, Minato-ku, Tokyo 106–8620, Japan) and a Resona 6 scanner (Mindray, Shenzhen, China) equipped with a L14-5WU transducer, respectively. The images were recorded by the Dezashib medical diagnostic imaging center (Tehran, Iran). This database comprised both standard mammographic (i.e., the CC and MLO views) and sonographic images from female patients, whose mean ages were 48.98 years, with a total of 156 solid masses, of which 77 and 79 were malignant and benign, respectively. Of these 156 masses, imaging follow-up of 39 masses (25%) remained stable for two or more years and thus, the final assessment was changed from BI-RADS 3 (probably benign) to BI-RADS 2 (benign). The benignancy or malignancy of other 117 masses (75%) was determined by biopsy. The mean image sizes of the mammograms and sonograms were 1504×1644 pixels and 910×1260 pixels, respectively. The spatial resolution in digital mammography was 50 microns/pixel. The X-ray tube voltage of 27 to 32 kVp was used to obtain these mammograms. The percentage frequency distribution of masses evaluated during this study is shown in Fig. 2. The mean sizes of malignant and benign masses were 20.50 mm and 20.26 mm, respectively.

A direct comparison of the reported results of different methods will not be fair enough when using different datasets. However, to the best of our knowledge, there are no publicly available datasets, which comprise of both standard mammographic and sonographic images from each female patient with the breast mass. Therefore, to deal with the absence of the public dataset in this field, previous studies were based on collecting and analyzing their own private datasets [2,11,34]. Thus, in

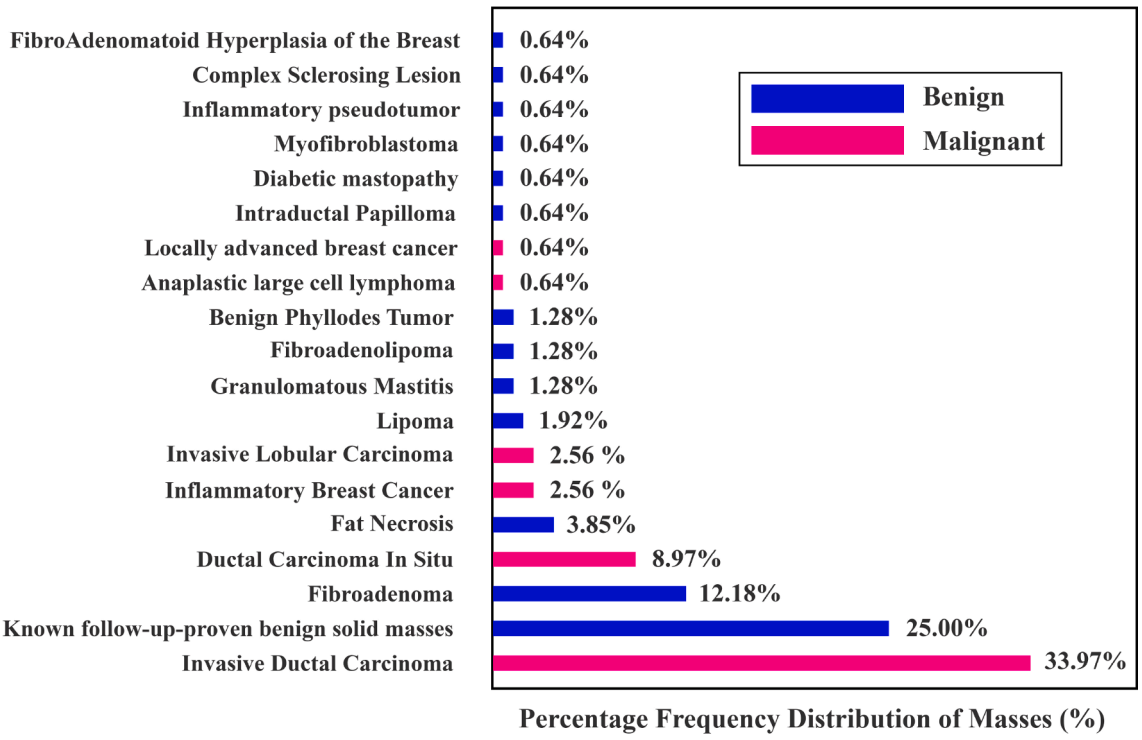

Fig. 2. Bar graph showing the percentage frequency distribution of masses evaluated during this study.

order to investigate if the proposed model can enhance the discrimination capability of benign and malignant solid breast masses in comparison with representative CAD systems in the literature, these systems were implemented and evaluated on our own dataset.

2.2. Data preprocessing

In computer vision, many tasks require image segmentation as a fundamental step. It allows the CAD system to best focus attention on the interested area and achieve better accuracy. This step is extremely important as it could influence the classification performance. In fact, the classification model relies on an accurate segmentation and could be sensitive to segmentation errors [35,36]. Therefore, on one hand, precise segmentations of mass regions from all three CC, MLO, and US images were necessary, in our study. On the other hand, it should be borne in mind that because multimodal analyses are usually used for further evaluation of difficult diagnostic cases, they are also usually 'challenging-to-segment' masses. Thus, the suspicious and complex appearance of these lesions makes the segmentation step much more difficult. Therefore, the current lack of reliable and validated (semi-) automatic segmentation software tools in such cases, demonstrate that manual segmentation remains important. For these reasons, in this study, to increase the radiologist's control over the final diagnosis, the segmentation stage was performed manually.

First, for each image (the CC, MLO, and US images), the expert delineated the contour of the lesion, manually. Second, a sub-image from every image was cropped by figuring out the smallest square containing the mass region and afterwards resized it into a fixed size (100×100). Third, the mass region was separated from the other tissue structures of every ROI using its binary mask and then fed into the model.

2.3. Data augmentation

In order to increase the number of datasets and reduce overfitting, artificial data augmentation is a solution widely used in the context of CNN. Therefore, the data augmentation strategy was applied to enlarge the dataset. To this end, we used three common augmentation operators as follows: (1) Flipping in vertical and horizontal directions; (2) Adding the random zero mean Gaussian white noise by randomly sampling variances from the interval [0.0005, 0.002] (twofold); (3) Rotating by randomly selecting angles from the interval (0^o, 360^o) (fourfold). As a result of this process, the dataset was increased by 24 times (i.e., 156 masses $\times$ 3 images per mass (the US, CC, and MLO images) $\times$ ($3 \times 2 \times 4$) $=$ 468 images $\times$ 24 $=$ 11,232 images).

2.4. Initialization of the proposed bimodal model using pre-trained mono-modal counterpart

Deep learning from scratch is very difficult due to the limited labeled training dataset in the medical field. Furthermore, it can be extremely time-consuming and tedious. However, a promising alternative to learning from scratch is to fine-tune a CNN that has been pre-trained using, for instance, a large set of labeled natural images [13]. Therefore, after fusion of the layers, additional convolutional layers are added to reduce the channel numbers so that the proposed bimodal model can also be initialized by the pre-trained mono-modal counterpart, as discussed in SubSection 2.5.

2.5. Proposed bimodal deep residual learning model

Although many investigations have demonstrated that network depth is a major factor of model expressiveness, the vanishing gradient, overfitting, and degradation problems make it difficult to train deeper networks effectively. Consequently, these problems have adverse impacts on classification. In the past several years, on the one hand, batch normalization is a widely used mechanism to alleviate the vanishing gradient problem. On the other hand, residual learning techniques are utilized to cope with the degradation and overfitting problems of deeper networks. These techniques allow the ResNet models to benefit the more network depth [14,15,37]. Considering that the deeper networks can acquire richer features for more accurate recognition, the ResNet models

achieved great successes in medical fields [7,13–15]. Therefore, in this paper, the ResNet with 50 layers [37] (ResNet50) is selected as the basic architecture of the proposed bimodal framework.

For the proposed model, six different configurations were considered. Then they extensively evaluated on our database to find the configuration with better performance for the proposed framework. How to design these configurations is shown in Fig. 3. In this regard, network architecture of the ResNet50 model is divided into five parts where each part is formed by approximately the similar layers. Afterward, the specific branches of the proposed model (i.e., the Branch1x for x = 1, 2, …, 6, Branch2y for y = 1, 2, 3, and Branch3), as shown in Fig. 4, are selected using the method is shown in Fig. 3. Furthermore, the architectures of these configurations are shown in Fig. 5, Table 1, and the codes are publicly available at github.com/Z-Assari/ResNet-Based-Models. The superscript of the $ResNet50_{i-j}$ expresses the layers of its branches, as shown in Fig. 3 and 5. The Branch1x, Branch2y, and Branch3 contain all layers between the start layers of the ResNet50 and Part i, the start layers of the Part i and Part j, and the start layer of the Part j to the last layer, respectively. In order to fuse the layers, they are concatenated in the third dimension. After fusion of the layers, additional convolutional layers are added to reduce the channel numbers so that the proposed bimodal model can also be initialized by the pre-trained mono-modal counterpart.

The superiority of the $ResNet50_{4-5}$ model compared with other configurations in all the evaluation metrics is shown in SubSection 3.1. The architecture of the $ResNet50_{4-5}$ model is illustrated in Fig. 5 (f). This model consists of the following four major steps. (1) First, the outputs of the seventh residual blocks are supposed as the high-level representations of the input images (i.e., the CC, MLO, and US images), in the aim to describe these images by significant feature maps. (2) Second, in order to explore complementary information between every two images effectively, these representations are concatenated in the third dimension, then additional 1 × 1 convolutional layers (in blue color) are added to reduce the channel numbers to 512. Afterward, the deep features are extracted by the next layers step by step to obtain the further identifying information from these fused feature maps. Then the outputs of the thirteenth blocks are supposed as the high-level joint representation of both images. (3) Third, all three of these representations (i.e., the joint representations of US and CC images, CC and MLO images, US and MLO images) are concatenated in the third dimension, afterwards additional 1 × 1 convolutional layer (in purple color) is added to reduce the channel number to 1024. Subsequently, the deep features are extracted by the later layers to provide much richer classification information from these fused feature maps. Then the output of the sixteenth block is supposed as the final common representation of the input images for the solid breast mass. (4) Eventually, the posterior probabilities for each class are assigned by the softmax classifier based on the information extracted from all input images. After completing the feature learning process, the final recognition result is obtained (i.e., whether this solid breast mass is benign or malignant). It is worth noting that our model, in accordance with the diagnostic procedure of radiologists previously mentioned, has a beneficial tendency to mammographic images, which influences the final recognition results. All networks were implemented with TensorFlow 1.12.0 Library in Python 3.6 on a desktop PC equipped with an Intel(R) Core(TM) i7-8086 K 4.00 GHz CPU with 16 GB of RAM, and an NVIDIA GeForce GTX 1080 Ti GPU.

2.6. Evaluation procedure

Five evaluation metrics -sensitivity (SE), specificity (SP), F1-score (F1), accuracy (AC), and area under ROC curve (AUC)- were adopted. The k-fold cross-validation is a very powerful tool to give us more accurate estimate of a model's performance. Therefore, in this study, the 5-fold cross-validation was used to evaluate the models more accurately [17]. The dataset was randomly divided into five subsets where each subset was formed by approximately the same proportions of the two types of class labels. Of these five folds, one fold is kept for testing, and the model is trained on remaining four augmented folds. This process is then repeated five times, with each of the five folds used exactly once as the test data. The results were evaluated as an average of the 5-fold cross-validation results. It should be noticed that no augmentation was performed on the test set. In this study, all methods were evaluated and optimized the same way.

3. Results

3.1. Comparison of the proposed bimodal configurations with their corresponding mono-modal counterparts

How to design the proposed bimodal configurations is discussed in SubSection 2.5. Results of mass classification for these models are shown in Table 2. Results demonstrate that best recognition results in all the metrics are achieved by the $ResNet50_{4-5}$ model.

3.2. The impact evaluation of the segmentation on the results

To this end, the lesion segmentation was again performed by another expert. Results demonstrate that the best recognition results on the SE, SP, F1, AUC, and AC metrics of 0.896, 0.910, 0.903, 0.965, and 0.903, respectively, are still achieved by the $ResNet50_{4-5}$ model. Fig. 6

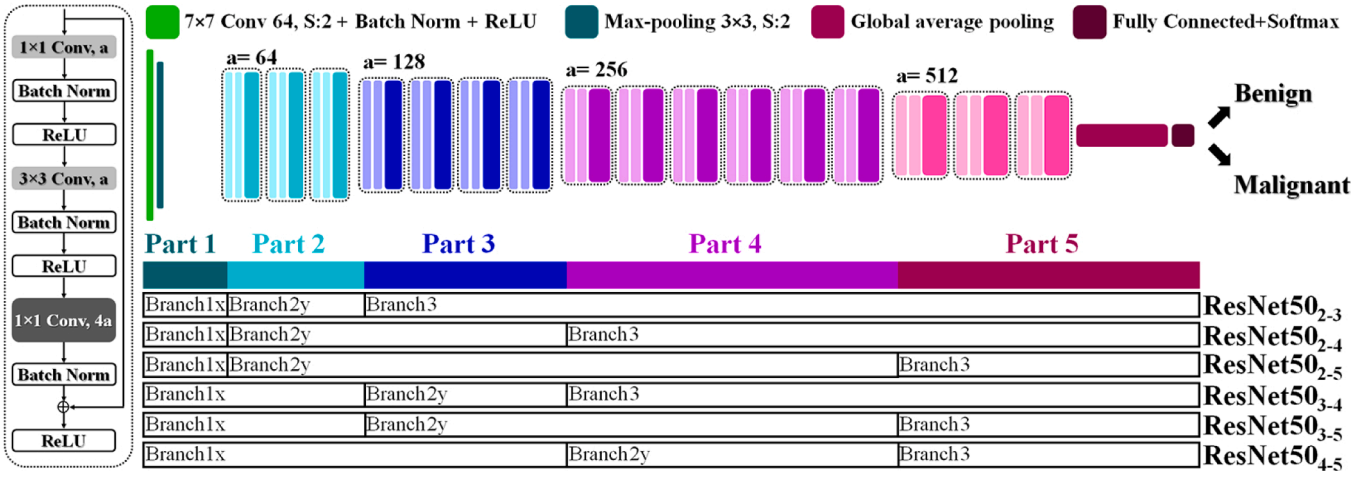

Fig. 3. The network architecture of the ResNet50 model used in the proposed bimodal framework. The block on the left of the ResNet50 model shows the architecture of a residual block.

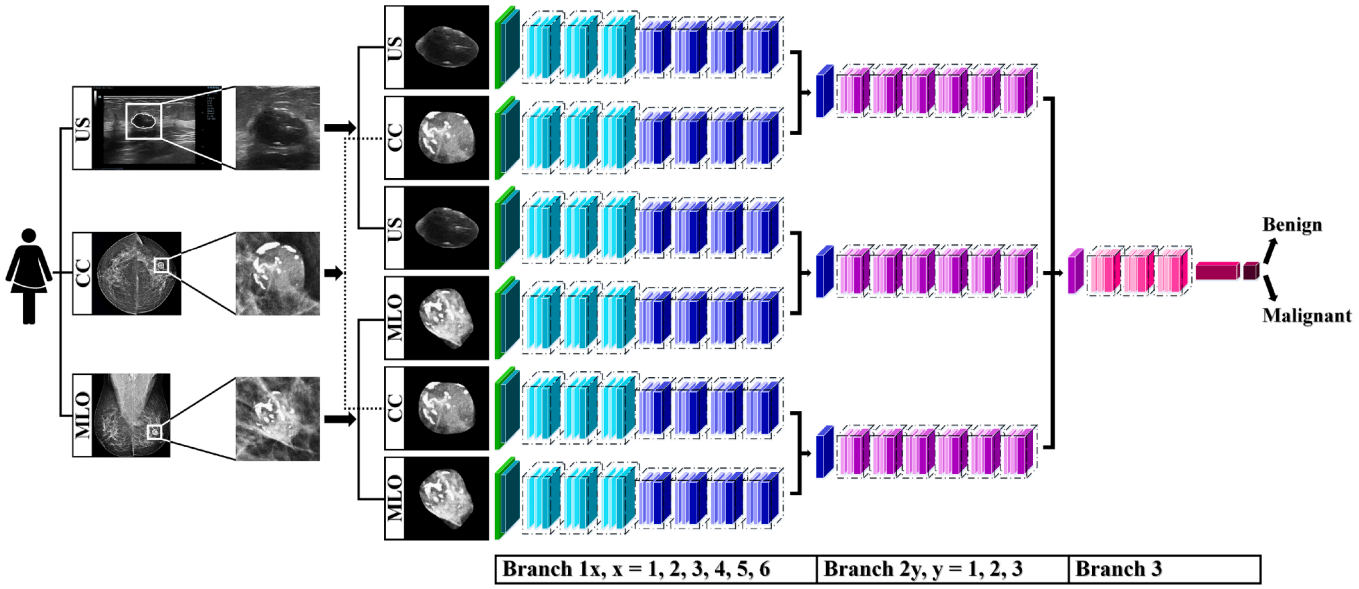

Fig. 4. Overall schematic of the proposed bimodal deep residual learning model.

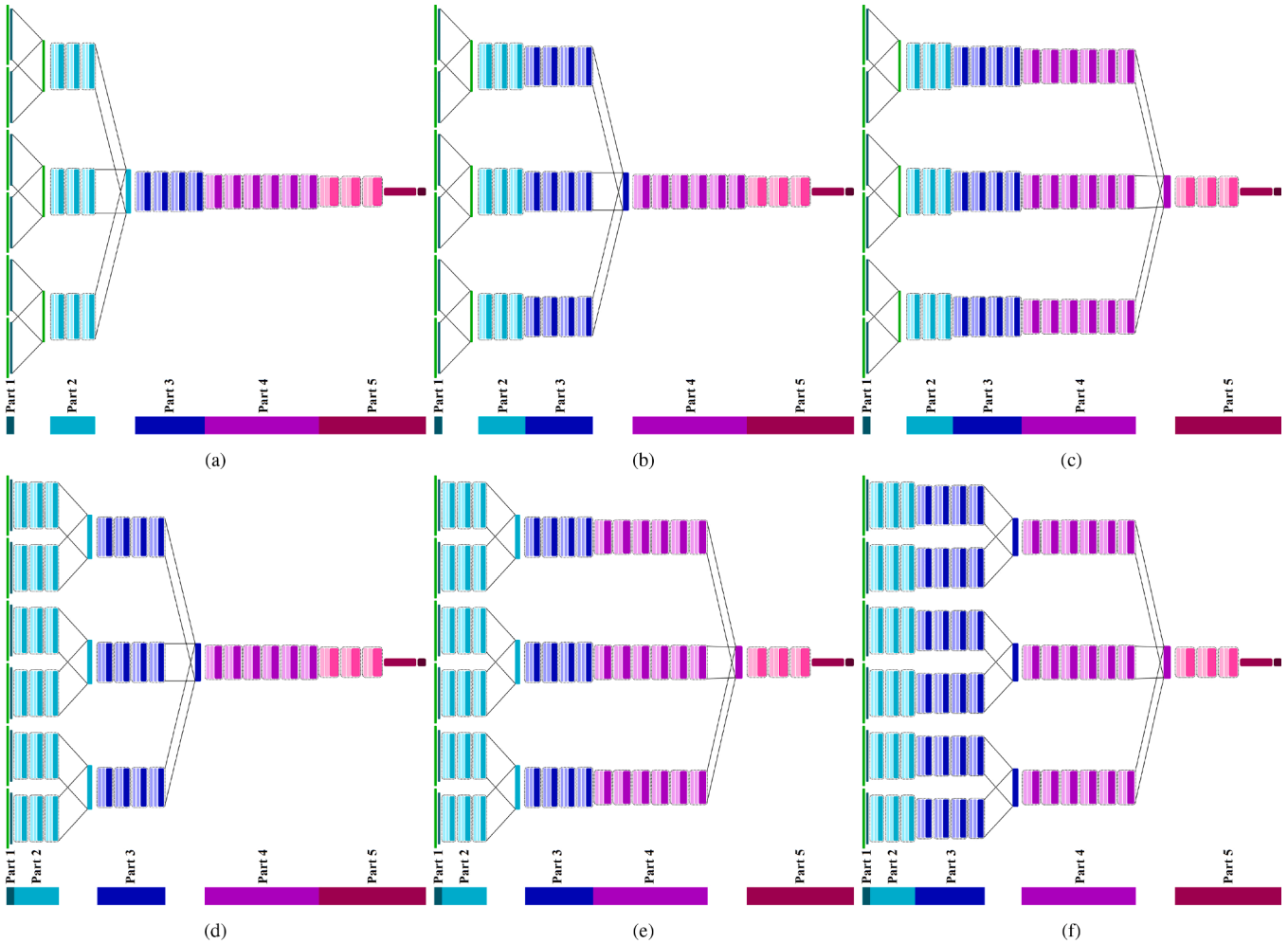

Fig. 5. The network architectures of the six different configurations of the proposed bimodal model (i.e., the (a) ResNet50$_{2-3}$, (b) ResNet50$_{2-4}$, (c) ResNet50$_{2-5}$, (d) ResNet50$_{3-4}$, (e) ResNet50$_{3-5}$, and (f) ResNet50$_{4-5}$ models).

Table 1
The network architectures of the ResNet50 model [37] and our six proposed bimodal configurations. Residual blocks are shown in brackets (see also Fig. 3, Fig. 4, and Fig. 5), with the numbers of blocks stacked. Conv(i × i,m,s): i × i convolutional layer with m filters and a stride of s. MaxPool(j × j,s): j × j max-pooling layer with a stride of s. FC(2): a fully connected layer with two units. AvgPool denotes a global average pooling layer. Concatenation denotes concatenating feature maps in third dimension. After fusion of the layers, additional convolutional layers are added to reduce the channel numbers.

ResNet50	$ResNet50_{2-3}$	$ResNet50_{2-4}$	$ResNet50_{2-5}$	$ResNet50_{3-4}$	$ResNet50_{3-5}$	$ResNet50_{4-5}$
Part 1: Conv(7 × 7,64,2) BatchNorm + Relu MaxPool(3 × 3,2)	**Branch 1x:** Conv(7 × 7,64,2) BatchNorm + Relu MaxPool(3 × 3,2) **Concatenation** Conv(7 × 7,64,1)	**Branch 1x:** Conv(7 × 7,64,2) BatchNorm + Relu MaxPool(3 × 3,2) **Concatenation** Conv(7 × 7,64,1)	**Branch 1x:** Conv(7 × 7,64,2) BatchNorm + Relu MaxPool(3 × 3,2) **Concatenation** Conv(7 × 7,64,1)	**Branch 1x:** Conv(7 × 7,64,2) BatchNorm + Relu MaxPool(3 × 3,2)	**Branch 1x:** Conv(7 × 7,64,2) BatchNorm + Relu MaxPool(3 × 3,2)	**Branch 1x:** Conv(7 × 7,64,2) BatchNorm + Relu MaxPool(3 × 3,2)
Part 2: $\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$	**Branch 2y:** $\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$ **Concatenation** Conv(1 × 1,256,1)	**Branch 2y:** $\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$	**Branch 2y:** $\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$ **Concatenation** Conv(1 × 1,256,1)	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$ **Concatenation** Conv(1 × 1,256,1)	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$
Part 3: $\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$	**Branch 3:** $\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$ **Concatenation** Conv(1 × 1,512,1)	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$	**Branch 2y:** $\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$ **Concatenation** Conv(1 × 1,512,1)	**Branch 2y:** $\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$ **Concatenation** Conv(1 × 1,512,1)
Part 4: $\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$	**Branch 3:** $\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$ **Concatenation** Conv(1 × 1,1024,1)	**Branch 3:** $\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$ **Concatenation** Conv(1 × 1,1024,1)	**Branch 2y:** $\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$ **Concatenation** Conv(1 × 1,1024,1)
Part 5: $\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$ AvgPool FC(2)+Softmax	$\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$ AvgPool FC(2)+Softmax	$\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$ AvgPool FC(2)+Softmax	**Branch 3:** $\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$ AvgPool FC(2)+Softmax	$\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$ AvgPool FC(2)+Softmax	**Branch 3:** $\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$ AvgPool FC(2)+Softmax	**Branch 3:** $\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$ AvgPool FC(2)+Softmax

Table 2
Results of mass classification for the proposed bimodal configurations:

Networks	SE	SP	F1	AUC	AC
$ResNet50_{US}$	0.845	0.848	0.844	0.925	0.846
$ResNet50_{CC}$	0.860	0.888	0.869	0.952	0.873
$ResNet50_{MLO}$	0.807	0.859	0.827	0.931	0.833
$ResNet50_{2-3}$	0.792	0.798	0.792	0.879	0.795
$ResNet50_{2-4}$	0.847	0.860	0.849	0.906	0.853
$ResNet50_{2-5}$	0.830	0.862	0.839	0.923	0.846
$ResNet50_{3-4}$	0.806	0.898	0.842	0.915	0.852
$ResNet50_{3-5}$	0.820	0.912	0.860	0.924	0.866
$ResNet50_{4-5}$	**0.898**	**0.938**	**0.916**	**0.964**	**0.917**

provides a glance at the confusion matrices of the obtained results of the proposed $ResNet50_{4-5}$ model with two different expert segmentations in the 5-fold cross validation.

3.3. Comparison with hand-crafted features based bimodal CAD systems

In order to investigate if the proposed deep learning-based model can enhance the discrimination capability of benign and malignant solid breast masses in comparison with representative hand-crafted features based bimodal CAD systems in the literature, their performances were compared with each other. For this comparison, these systems were

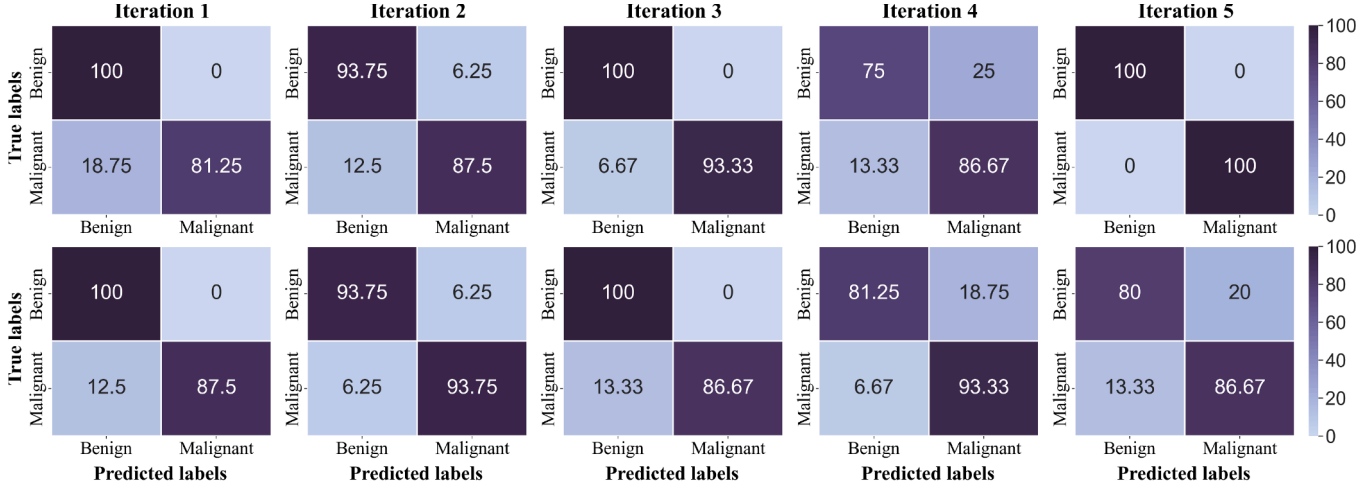

Fig. 6. Confusion matrices of the obtained results of the proposed $ResNet50_{4-5}$ model with two different expert segmentations in the 5-fold cross validation. Each row represents the results that are achieved with one expert segmentation. These results have been expressed as percentages.

Table 3
Comparison with hand-crafted features based bimodal CAD systems:

Networks	SE	SP	F1	AUC	AC
Drukker et al. [2]	0.600	0.804	0.598	0.727	0.717
Sidiropoulos et al. [11]	0.667	0.625	0.615	0.646	0.643
$ResNet50_{4-5}$	**0.898**	**0.938**	**0.916**	**0.964**	**0.917**

implemented and evaluated on the dataset of this paper. It can be seen from Table 3 that our model achieved far much better performance compared with them.

3.4. Comparison with deep learning-based CAD systems

We compared our model, the $ResNet50_{4-5}$, with several recently published deep learning-based models for breast cancer diagnosis from mammography and/or US images. These models were proposed by Lévy & Jain [26] (2016), Han et al. [28] (2017), Moon et al. [32] (2020), and Habib et al. [34] (2020). All these models have reported improvements over existing state-of-the-art results. For this comparison, these CAD systems were implemented and evaluated on the dataset of this paper. The results and detailed information about these models are shown in Table 4. It can be seen from Table 4 that best recognition results on the SE, SP, F1, AUC, and AC metrics of 0.898, 0.938, 0.916, 0.964, and 0.917, respectively, are achieved by the $ResNet50_{4-5}$ model.

3.5. Model generalization

Generalization is a term used to describe a model's ability to perform on the test dataset. In the machine learning field, overfitting is a very important and closely related topic to generalization. Overfitting is a modeling error that occurs when the model performs well on the training data and poorly on the test data. Overfitting can be identified by monitoring the performance of the model during training phase by evaluating it on both the training and test datasets [37]. In Fig. 7, the mean accuracy curve plots of the proposed bimodal model and its corresponding mono-modal counterparts on the training and test datasets are plotted together, during training procedure. We have two major observations from this figure. First, Fig. 7 shows that even until the ten thousandth epoch, the $ResNet50_{4-5}$ model does not begin to overfit, just as well as its corresponding mono-modal counterparts. Second, compared to the $ResNet50_{US}$, $ResNet50_{CC}$, and $ResNet50_{MLO}$ models, the $ResNet50_{4-5}$ model has higher testing accuracy throughout the whole training procedure. This comparison verifies the effectiveness of the proposed bimodal model in solid breast mass classification. Thus, the data augmentation and residual learning techniques used in this study not only avoid overfitting, but they also lead to better overall model performance.

3.6. Qualitative evaluation

Fig. 8 (a) and (b) demonstrate the ability of our model to improve the solid breast mass recognition. These challenging cases are with both benign and malignant features that make it difficult to distinguish them using information from images of a single modality. Therefore, they are frequently determined as suspicious by the mono-modal models. Although our model can distinguish malignant masses from benign ones on the various challenging conditions accurately, there are still some failure cases, as shown in Fig. 8 (c) and (d). The majority of misclassification cases are calcified. That makes sense, because microcalcification suggest a mass is more likely to be malignant.

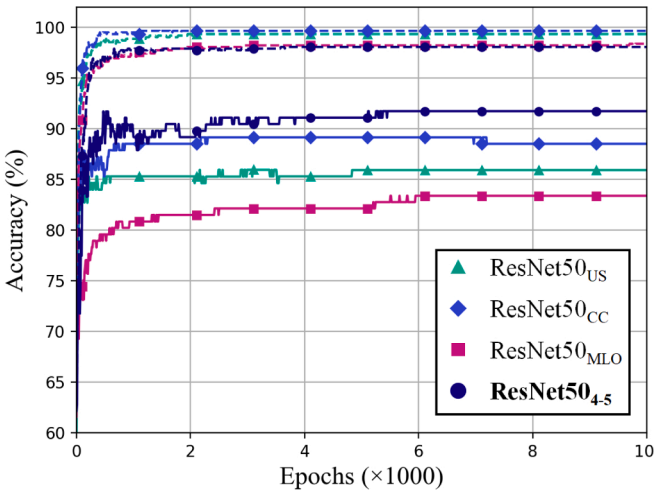

Fig. 7. The mean accuracy curve plots on the training (dashed lines) and test datasets (bold lines), during model training in the 5-fold cross-validation scenario.

4. Discussion

The superiority of the $ResNet50_{4-5}$ model compared with other configurations and its corresponding mono-modal counterparts in all the evaluation metrics demonstrates that higher-level feature maps are more suitable for feature-maps fusion. It can be seen from Table 2 that best recognition results on the SE, SP, F1, AUC, and AC metrics of 0.898, 0.938, 0.916, 0.964, and 0.917, respectively, are achieved by the $ResNet50_{4-5}$ model. Therefore, it can be concluded from these experiments when the information obtained from images of a single modality is inadequate and the masses determined as suspicious, this model can effectively leverage the complementary information from two modalities to enhance the solid mass classification.

The performances of the proposed model and representative hand-crafted features based bimodal CAD systems in the literature were compared with each other. Drukker et al. [2] developed a CAD system to distinguish malignant lesions from benign ones. Lesion classification was performed by a Bayesian neural network with five units in a single hidden layer. Results showed a statistically significant increase over that

Table 4
Comparison with deep learning-based CAD systems.

Reference	Modality	Based on	SE	SP	F1	AUC	AC
Lévy & Jain [26]	CC	Shallow CNN	0.792	0.708	0.755	0.838	0.750
	CC	AlexNet	0.733	0.938	0.799	0.942	0.835
	CC	GoogleNet	0.859	0.899	0.875	0.959	0.879
	MLO	Shallow CNN	0.721	0.592	0.592	0.751	0.655
	MLO	AlexNet	0.765	0.812	0.765	0.868	0.787
	MLO	GoogleNet	0.832	0.872	0.849	0.927	0.853
Han et al. [28]	US	GoogleNet	0.883	0.782	0.843	0.905	0.833
Moon et al. [32]	US	DenseNet	0.807	0.847	0.822	0.905	0.827
Habib et al. [34]	US + CC	GoogleNet	0.874	0.911	0.884	0.949	0.892
	US + MLO	GoogleNet	0.871	0.897	0.882	0.945	0.885
Our model	US + CC + MLO	ResNet50	**0.898**	**0.938**	**0.916**	**0.964**	**0.917**

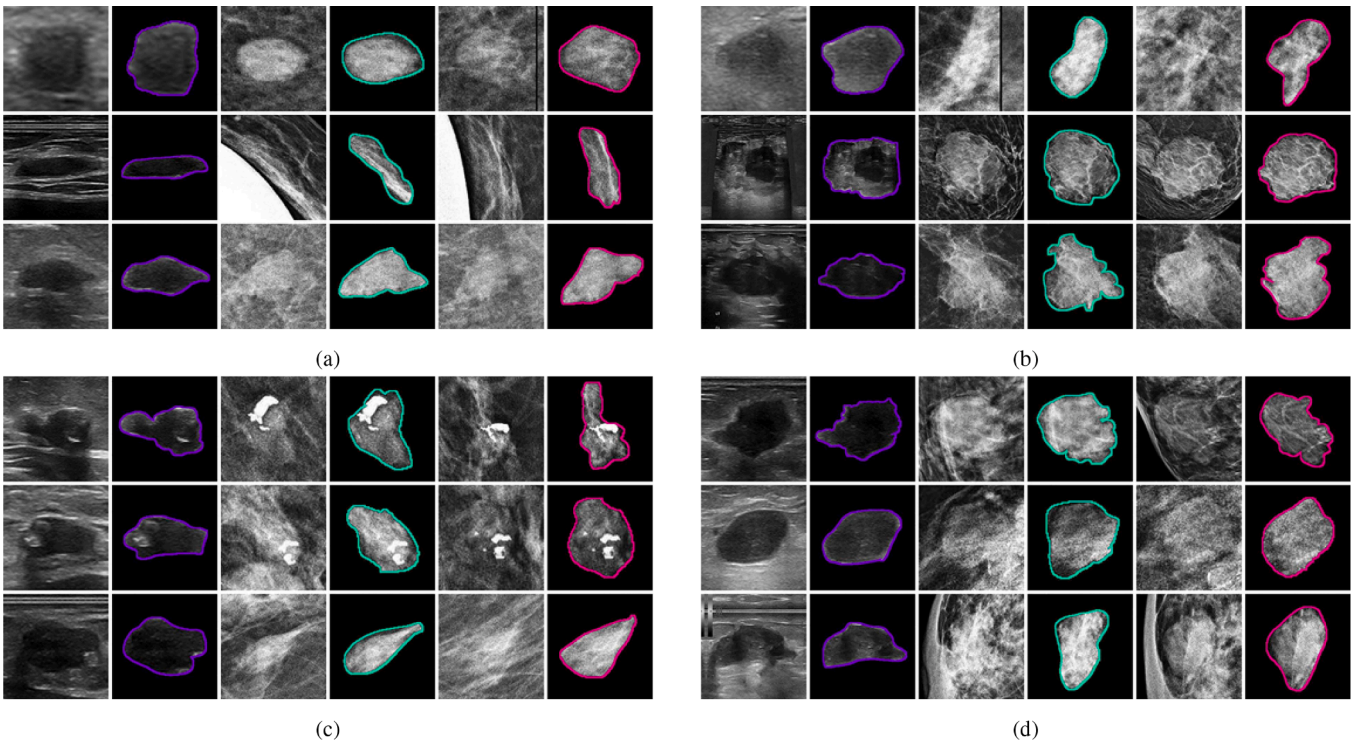

Fig. 8. Examples of some benign and malignant solid breast masses correctly classified or misclassified by the proposed model. Each row represents one mass and from left to right are US ROI, segmented US ROI, CC ROI, segmented CC ROI, MLO ROI, and segmented MLO ROI, respectively. (a) True Negatives, (b) True Positives, (c) False Positives, and (d) False Negatives.

achieved with features from either mammography or ultrasound alone. By evaluating this model on our dataset, the recognition results on the SE, SP, F1, AUC, and AC metrics of 0.600, 0.804, 0.598, 0.727, and 0.717, respectively, were obtained, as shown in Table 3.

Sidiropoulos et al. [11] designed a GPU-based CAD system to accurately discriminate benign to malignant lesions. They employed the probabilistic neural network classifier and exhaustive features selection method to design the system. Therefore, they utilized the parallel processing programming technique on the processors of the GPU to deal with the enormous computational load of searching exhaustively for best feature combinations in a large feature space. By evaluating this model on our dataset, the recognition results on the SE, SP, F1, AUC, and AC metrics of 0.667, 0.625, 0.615, 0.646, and 0.643, respectively, were obtained, as shown in Table 3. These experimental results demonstrate the superiority of our model compared with them. Thus, due to exploit the high-level features, our deep learning-based CAD system achieved superior performance compared with hand-crafted features based ones.

Our model, the ResNet50$_{4-5}$, was compared with several recently published deep learning-based models for breast cancer diagnosis from mammography and/or US images. These models were proposed by Lévy & Jain [26] (2016), Han et al. [28] (2017), Moon et al. [32] (2020), and Habib et al. [34] (2020). Results of mass classification for these models are shown in Table 4. Habib et al. [34] developed a bimodal GoogleNet-based CAD system to distinguish malignant breast lesions from benign ones. The experimental results showed improvement in AUC when utilizing both modalities. There are five main differences between this model and the ResNet50$_{4-5}$ model. In our study: (1) The ResNet50 model selected as the basic architecture of the bimodal framework; (2) The feature-maps fusion, instead of combining the vectors, is used to implement combinations of the images' features of different modalities; (3) Inspired by the approach of radiologists, the two standard mammographic views (i.e., the CC and MLO images) and US images are considered as separate inputs; (4) The proposed model is trained for these three images at the same time; (5) This study focuses on a specific type of mass (i.e., solid mass). It can be seen from Table 4 that best recognition results in all the metrics are achieved by the ResNet50$_{4-5}$ model.

In this study, the high-level representations inspired by the approach of some radiologists during the radiologic examination using the images of different imaging modalities were combined. (1) First, they focus attention on the interested area of each image separately. (2) Afterward, mass characteristics of every two images are considered, with the aim of obtaining the further diagnostic information. (3) Finally, the diagnostic decision is obtained based on information extracted from all images. Therefore, the proposed bimodal model extracts more informative features from input images and achieves superior performance, as the results of extensive experiments of this study show.

Several limitations of this study and potential ways to overcome them can be considered. (1) There exist many publicly available databases to design mono-modal CAD systems in this field. However, in order to investigate a bimodal CAD system, it is essential that the dataset contains the sonographic and mammographic images collected from each patient. Although our dataset size is in the range with previous works in the bimodal classification literature (Drukker et al. [2], 100 lesions; Sidiropoulos et al. [11], 62 lesions; and Habib et al. [34], 153 lesions), a further evaluation would require a larger dataset. (2) The cases collected for this study were not sufficiently diverse. To verify the applicability of a model, various types of cancers should be included in further investigation. (3) Most of the cases that were incorrectly classified by our model were calcified masses. Considering that mass characteristics such as the margin, shape, density, orientation, echo pattern, and posterior acoustic features, contain valuable recognition information [9], they can be analyzed to extract more descriptive features that can be used in deep learning-based CAD systems. (4) Because multimodal analyses are usually used for further evaluation of difficult diagnostic cases, the manual segmentation increases the radiologist's control over the final diagnosis. Nevertheless, this human intervention limits the model to be part of a fully automatic system.

5. Conclusion

In this paper, a novel bimodal deep residual learning model is presented to meet the challenges of the information combination from mammography and US imaging in solid breast mass classification. To the best of our knowledge, the proposed framework is among the earliest efforts of applying deep learning approaches to design bimodal CAD system. It is worthwhile to point out that our framework is general enough that can be easily extended to other complicated medical image analysis tasks sharing the similar challenges of bimodal suspicious mass analysis. Extensive experiments demonstrate the effectiveness and superiority of the proposed bimodal framework for solid breast mass classification compared with other state-of-the-art models. As a part of our future studies, a reader study involving a group of radiologists is needed to prove that the proposed model leads to clinical benefits. In addition, it is proposed to design multimodal CAD Systems using various fusion strategies to implement combinations of the images' features of different modalities properly for more accurate classification in medical image analysis tasks.

Ethical approval

Approval was obtained from the ethics committee of Tarbiat Modares University. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

CRediT authorship contribution statement

Zahra Assari: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Investigation, Visualization, Writing - original draft. **Ali Mahloojifar:** Conceptualization, Methodology, Validation, Investigation, Project administration, Supervision, Writing - review & editing. **Nasrin Ahmadinejad:** Conceptualization, Validation, Investigation, Resources, Data curation.

Declaration of Competing Interest

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

References

[1] S. Perek, N. Kiryati, G. Zimmerman-Moreno, M. Sklair-Levy, E. Konen, A. Mayer, Classification of contrast-enhanced spectral mammography (CESM) images, Int. J. Comput. Assist. Radiol. Surg. 14 (2) (2019) 249–257, https://doi.org/10.1007/s11548-018-1876-6.

[2] K. Drukker, K. Horsch, M.L. Giger, Multimodality computerized diagnosis of breast lesions using mammography and sonography1, Acad. Radiol. 12 (8) (2005) 970–979, https://doi.org/10.1016/j.acra.2005.04.014.

[3] Y.-H. Yu, W. Wei, J.-L. Liu, Diagnostic value of fine-needle aspiration biopsy for breast mass: a systematic review and meta-analysis, BMC Cancer 12 (1) (2012) 1–14, https://doi.org/10.1186/1471-2407-12-41.

[4] O. Graf, T.H. Helbich, G. Hopf, C. Graf, E.A. Sickles, Probably benign breast masses at US: is follow-up an acceptable alternative to biopsy? Radiology 244 (1) (2007) 87–93, https://doi.org/10.1148/radiol.2441060258.

[5] M.B. Mainiero, A. Goldkamp, E. Lazarus, L. Livingston, S.L. Koelliker, B. Schepps, W.W. Mayo-Smith, Characterization of breast masses with sonography: can biopsy of some solid masses be deferred? J. Ultrasound Med. 24 (2) (2005) 161–167, https://doi.org/10.7863/jum.2005.24.2.161.

[6] O. Graf, T.H. Helbich, M.H. Fuchsjaeger, G. Hopf, M. Morgun, C. Graf, R. Mallek, E. A. Sickles, Follow-up of palpable circumscribed noncalcified solid breast masses at mammography and US: can biopsy be averted? Radiology 233 (3) (2004) 850–856, https://doi.org/10.1148/radiol.2333031845.

[7] Z. Jiao, X. Gao, Y. Wang, J. Li, A parasitic metric learning net for breast mass classification based on mammography, Pattern Recognit. 75 (2018) 292–301, https://doi.org/10.1016/j.patcog.2017.07.008.

[8] B. Wilczek, H.E. Wilczek, L. Rasouliyan, K. Leifland, Adding 3D automated breast ultrasound to mammography screening in women with heterogeneously and extremely dense breasts: report from a hospital-based, high-volume, single-center breast cancer screening program, Eur. J. Radiol. 85 (9) (2016) 1554–1563, https://doi.org/10.1016/j.ejrad.2016.06.004.

[9] A.C. of Radiology, C.J. D'Orsi, ACR BI-RADS Atlas: Breast Imaging Reporting and Data System; Mammography, Ultrasound, Magnetic Resonance Imaging, Follow-up and Outcome Monitoring, Data Dictionary, ACR, American College of Radiology, 2013.

[10] B. Sahiner, H.-P. Chan, L.M. Hadjiiski, M.A. Roubidoux, C. Paramagul, J.E. Bailey, A.V. Nees, C.E. Blane, D.D. Adler, S.K. Patterson, Multi-modality CADx: ROC study of the effect on radiologists' accuracy in characterizing breast masses on mammograms and 3D ultrasound images, Acad. Radiol. 16 (7) (2009) 810–818, https://doi.org/10.1016/j.acra.2009.01.011.

[11] K.P. Sidiropoulos, S.A. Kostopoulos, D.T. Glotsos, E.I. Athanasiadis, N. D. Dimitropoulos, J.T. Stonham, D.A. Cavouras, Multimodality GPU-based computer-assisted diagnosis of breast cancer using ultrasound and digital mammography images, Int. J. Comput. Assist. Radiol. Surg. 8 (4) (2013) 547–560, https://doi.org/10.1007/s11548-013-0813-y.

[12] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, R. Zwiggelaar, Deep learning in mammography and breast histology, an overview and future trends, Med. Image Anal. 47 (2018) 45–67, https://doi.org/10.1016/j.media.2018.03.006.

[13] J. Han, H. Chen, N. Liu, C. Yan, X. Li, CNNs-based rgb-d saliency detection via cross-view transfer and multiview fusion, IEEE T. Cybern. 48 (11) (2017) 3171–3183, https://doi.org/10.1109/TCYB.2017.2761775.

[14] H. Chen, X. Yuan, Z. Pei, M. Li, J. Li, Triple-classification of respiratory sounds using optimized s-transform and deep residual networks, IEEE Access 7 (2019) 32845–32852, https://doi.org/10.1109/ACCESS.2019.2903859.

[15] L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. Med. Imaging 36 (4) (2016) 994–1004, https://doi.org/10.1109/TMI.2016.2642839.

[16] L. Shen, M. He, N. Shen, N. Yousefi, C. Wang, G. Liu, Optimal breast tumor diagnosis using discrete wavelet transform and deep belief network based on improved sunflower optimization method, Biomed. Signal Process. Control 60 (2020), 101953, https://doi.org/10.1016/j.bspc.2020.101953.

[17] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. Van Riel, M.M.W. Wille, M. Naqibullah, C.I. S'anchez, B. van Ginneken, Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks, IEEE Trans. Med. Imaging 35 (5) (2016) 1160–1169, https://doi.org/10.1109/TMI.2016.2536809.

[18] P. Wang, Q. Song, Y. Li, S. Lv, J. Wang, L. Li, H. Zhang, Cross-task extreme learning machine for breast cancer image classification with deep convolutional features, Biomed. Signal Process. Control 57 (2020), 101789, https://doi.org/10.1016/j.bspc.2019.101789.

[19] P. Wang, J. Wang, Y. Li, P. Li, L. Li, M. Jiang, Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing, Biomed. Signal Process. Control 65 (2021), 102341, https://doi.org/10.1016/j.bspc.2020.102341.

[20] O.N. Oyelade, A.E. Ezugwu, A deep learning model using data augmentation for detection of architectural distortion in whole and patches of images, Biomed. Signal Process. Control 65 (2021), 102366, https://doi.org/10.1016/j.bspc.2020.102366.

[21] D.R. Nayak, R. Dash, B. Majhi, R.B. Pachori, Y. Zhang, A deep stacked random vector functional link network autoencoder for diagnosis of brain abnormalities and breast cancer, Biomed. Signal Process. Control 58 (2020), 101860, https://doi.org/10.1016/j.bspc.2020.101860.

[22] G. Carneiro, J. Nascimento, A.P. Bradley, Unregistered multiview mammogram analysis with pre-trained deep learning models, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 652–660, doi: 10.1007/978-3-319-24574-478.

[23] J. Arevalo, F.A. Gonz'alez, R. Ramos-Poll'an, J.L. Oliveira, M.A.G. Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, Comput. Meth. Prog. Bio. 127 (2016) 248–257, https://doi.org/10.1016/j.cmpb.2015.12.014.

[24] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, C.-M. Chen, Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans, Sci. Rep. 6 (1) (2016) 1–13, https://doi.org/10.1038/srep24454.

[25] B.Q. Huynh, H. Li, M.L. Giger, Digital mammographic tumor classification using transfer learning from deep convolutional neural networks, J. Med. Imaging 3 (3) (2016), 034501, https://doi.org/10.1117/1.JMI.3.3.034501.

[26] D. Lévy, A. Jain, Breast mass classification from mammograms using deep convolutional neural networks, arXiv preprint arXiv:1612.00542 (2016).

[27] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, L. Li, Discrimination of breast cancer with microcalcifications on mammography by deep learning, Sci. Rep. 6 (1) (2016) 1–9, https://doi.org/10.1038/srep27327.

[28] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, Y.-K. Seong, A deep learning framework for supporting the classification of breast lesions in ultrasound images, Phys. Med. Biol. 62 (19) (2017) 7714, https://doi.org/10.1088/1361-6560/aa82ec.

[29] W. Sun, T.-L.B. Tseng, J. Zhang, W. Qian, Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data, Comput. Med. Imaging Graph. 57 (2017) 4–9, https://doi.org/10.1016/j.compmedimag.2016.07.004.

[30] L. Shen, L.R. Margolies, J.H. Rothstein, E. Fluder, R. McBride, W. Sieh, Deep learning to improve breast cancer detection on screening mammography, Sci. Rep. 9 (1) (2019) 1–12, https://doi.org/10.1038/s41598-019-48995-4.

[31] X. Yu, N. Zeng, S. Liu, Y.-D. Zhang, Utilization of DenseNet201 for diagnosis of breast abnormality, Mach. Vis. Appl. 30 (7) (2019) 1135–1144, https://doi.org/10.1007/s00138-019-01042-8.

[32] W.K. Moon, Y.-W. Lee, H.-H. Ke, S.H. Lee, C.-S. Huang, R.-F. Chang, Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks, Comput. Meth. Prog. Bio. 190 (2020), 105361, https://doi.org/10.1016/j.cmpb.2020.105361.

[33] Y.-D. Zhang, S.C. Satapathy, D.S. Guttery, J.M. G'orriz, S.-H. Wang, Improved breast cancer classification through combining graph convolutional network and convolutional neural network, Inform. Process. Manage. 58 (2) (2021), 102439, https://doi.org/10.1016/j.ipm.2020.102439.

[34] G. Habib, N. Kiryati, M. Sklair-Levy, A. Shalmon, O.H. Neiman, R.F. Weidenfeld, Y. Yagil, E. Konen, A. Mayer, Automatic breast lesion classification by joint neural analysis of mammography and ultrasound, in: Multimodal Learning for Clinical Decision Support and Clinical Image-Based Procedures: 10th international workshop, ML-CDS 2020, and 9th international workshop, CLIP 2020, held in conjunction with MICCAI 2020, Lima, Peru, October 4–8, Proceedings, Vol. 12445, Springer, 2020, pp. 125–135, doi: 10.1007/978-3-030-60946-7-13.

[35] K.B. Soulami, N. Kaabouch, M.N. Saidi, A. Tamtaoui, Breast cancer: one-stage automated detection, segmentation, and classification of digital mammograms using unet model based-semantic segmentation, Biomed. Signal Process. Control 66 (2021), 102481, https://doi.org/10.1016/j.bspc.2021.102481.

[36] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, M. Andre, Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network, Biomed. Signal Process. Control 61 (2020), 102027, https://doi.org/10.1016/j.bspc.2020.102027.

[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016, pp. 770–778.