Journal of King Saud University – Computer and Information Sciences 34 (2022) 3310–3318

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Feature fusion and Ensemble learning-based CNN model for mammographic image classification

Imran Ul Haq [a], Haider Ali [a], Hong Yu Wang [b], Cui Lei [a,*], Hazrat Ali [c,*]

[a] School of Information Science and Technology, Northwest University, Xi'an 710127, China
[b] School of Computer Science, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
[c] Department of Electrical and Computer Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad, Pakistan

ARTICLE INFO

Article history:
Received 29 November 2021
Revised 26 February 2022
Accepted 24 March 2022
Available online 19 April 2022

Keywords:
Breast cancer
Computer-Aided Diagnosis (CAD)
Feature fusion
Ensemble learning
Mammograms classification

ABSTRACT

In recent times, the world has faced an alarming situation regarding breast cancer patients. The early diagnosis of this deadly disease can make the treatment more accessible and practical. In this regard, a Computer-Aided Diagnosis (CAD) system can assist the radiologists in distinguishing the normal and abnormal tissues and diagnosing the pathological stages. The classification task is challenging in CAD systems because of noisy and low contrast mammogram images, tumors' shape and location variations, and the high resemblance between the normal and tumor regions of interest (ROI). We propose a novel deep convolution neural network (DCNN) approach based on feature fusion and ensemble learning strategies to improve the detection and classification of abnormalities in mammographic scans. The feature fusion helps to detect discriminative features between the classes properly, while ensemble learning in the last block better classify the normal and tumor ROIs to get more authenticated results. Moreover, the role of spatial dropout and depthwise separable convolution is investigated for mammogram classification to better deal with overfitting and small dataset problems in medical images. The proposed model is evaluated on two publicly available datasets, MIAS and BCDR, getting high sensitivity, specificity, and accuracy of 0.995, 0.994, and 0.994, respectively, on the MIAS dataset.

1. Introduction

Breast cancer is considered one of the leading cancers that cause the death of women worldwide. It is a source of 23% of cancer cases, which leads to 14% of deaths in women (Whang et al., 2013). According to research conducted by a leading breast cancer awareness organization (Hudson, 2019), about 12% of women, which means 1 out of 8 women may have breast cancer in any time frame of their life span. According to a 2022 study of the American Cancer Society (ACS), there would be an expected 287,850 new patients with invasive breast cancer diagnosed in females, 2,710 patients diagnosed in males, and 51,400 patients of ductal carcinoma in situ detection in women (Siegel et al., 2022). After lung cancer, breast cancer is the next highest cause of cancer mortality among women in the United States and, according to estimates, will kill about 43,250 women and 530 men in 2022 (Siegel et al., 2022). These statistics clearly depict the need for an efficient model for the early detection of this deadly disease. According to medical research, the early diagnosis of breast cancer can play an essential role in a patient's speedy recovery and reducing the mortality rate (Oeffinger et al., 2015). Therefore researchers are proposing new techniques for early diagnosis, especially in medicine and computational intelligence. Mammographic scans are commonly utilized as a source for early detection (Olsen and Gøtzsche, 2001). Although, many imaging techniques are being utilized for the diagnosis of breast cancer which may include Computed Tomography (CT) Imaging (Shi et al., 2019), ultrasonography (US) (Jalalian et al., 2013), and Magnetic Resonance Imaging (MRI) (Mann et al., 2008), mammograms have shown remarkable results to help reduce breast cancer risk. They have greatly benefited women patients aged between 40 and 49 years (Smart et al., 1995).

At the screening process, a mammogram image is examined by a radiologist. To ensure that the process remains error-free, the radiologist must be an expert in this field, and still, there is a

* Corresponding authors.
E-mail addresses: alihaider@stumail.nwu.edu.cn (H. Ali), hywang@xupt.edu.cn (H.Y. Wang), leicui@nwu.edu.cn (C. Lei), hazratali@cuiatd.edu.pk (H. Ali).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2022.03.023
1319-1578/

chance of human error. The manual examination of the mammogram images is an exhaustive job, and a small human error can result in the wrong diagnosis, resulting in incurable cancer patients. In this regard, CAD can play its role in facilitating radiologists in differentiating normal and abnormal scans for better diagnosis. Moreover, CAD can provide fast results which won't be prone to human error. However, the results recommended by the CAD system can also be re-examined by the radiologist to have a cross-check.

The CAD system comprises different stages that a mammogram image has to undergo to get the prediction. The typical stages may include preprocessing the input image, extraction of the regions of interest (ROI), feature extraction, and classification. This process is arduous because of noise in mammograms; different texture and shape details of masses. Blood arteries and muscular fibers further hinder accurate detection. Because of these constraints, it is pretty challenging to discover effective patterns. Many systems have been proposed in the literature containing traditional machine learning (ML) and deep neural networks to deal with these problems.

The traditional ML techniques use handcrafted features for classification (Tang et al., 2009; Wei et al., 2005; Wei et al., 2005). Oliver et al. (2012) proposed a model to extract local features using a set of high pass filters and select the most effective features to train AdaBoost SVM for classification of ROI into the tumor and normal. In another work presented by Khan et al. (2016), the author proposes a Gabor filter bank to optimize the detection of breast cancer in monograms. The proposed work was supported by an additive clustering algorithm and Particle Swarm Optimization (PSO). For PSO, the Gaussian kernel was used as a fitness function in the SVM. The proposed work was tested on the large number of mammographic images taken from the available online dataset to observe the implication of the optimized Gabor filter bank. For evaluation, statistical parameters like total classification accuracy, the area under the curve (AUC), sensitivity, and specificity were used. Arevalo et al. (2016) utilized learned and handcrafted representations of ROIs to train the hybrid classification model. Similarly, Jiao et al. (2016) proposed a feature vector consisting of learned features and intensity-based features. Moayedi et al. (2010) proposed a diagnostic classification system, where a feature vector was obtained from contourlet coefficients and the genetic algorithm to advance the assortment of extracted features. Lastly, an SVM was used for the final classification.

In ML-based techniques, support vector machine (SVM) and random forest (RF) are considered the most influential classifiers for mammogram classification. But in ML-based practices most tedious job is to extract an effective feature vector to train the model.

In recent years great development has been observed in artificial intelligence reason being the deep learning algorithms (Rehman and Chong, 2020). The best part of Deep Neural Network (DNN) is that it extracts the features on its own; therefore, we don't need to try optimal feature extractors (ur Rehman et al., 2018). The deep learning-based technique has solved the problem of attaining more information from the input data by intense learning and holds the capability of extracting and learning high-level features. In the DNN model, higher accuracy is achieved as an iterative process where a feature vector set along with weights are optimal to best fit. Moreover, the dropout layers in any DNN model exclude the weak features from the optimal feature vector set. Deep learning models have shown remarkable results in many fields (D'Angelo and Palmieri, 2021; D'Angelo and Palmieri, 2020), including medical imaging (Cai, 2021; Guo and Razmjooy, 2019; Liu, 2020; Razmjooy et al., 2020; Rehman, 2021).

Wang et al. (2016) developed a stacked denoising auto-encoder for the combined assessment of breast masses and microcalcifications. The model was based on deep learning to diagnose breast cancer and achieved good performance. A CAD system is proposed in Agarwal et al. (2019), where the transfer learning technique is utilized in the model to make the prediction. The author proposed a robust method for accurate classification of breast cells into two categories which are normal and abnormal. Huynh et al. (2016) investigated the use of deep learning frameworks to retrieve features of breast tumors and found that they performed well when evaluated to analytically retrieved features. Li et al. (Wang et al., 2016) used CNN to classify abnormalities in breast scans as malignant or benign and got high sensitivity. Cai et al. (2019) suggested a CNN model for diagnosing breast microcalcification. To fully exploit classical features' potential, they combined handcrafted and CNN features to increase the classification results. Lévy and Jain (1612) used a CNN to categorize the sampled breast tumors as malignant or benign and got good results. Zhu et al. (2017) developed a multi-instance deep learning model for masses classification using full mammography scans.

Our proposed CAD system uses a customized CNN model to classify the tumor and normal ROIs collected from mammograms. The proposed model first preprocesses the input raw image to remove noise and increase the quality. The second step extracts the ROIs of normal and abnormal classes. Next, a CNN model extracts features from the collected ROIs using feature fusion and give them to three different classifiers to decide the class of the ROI using majority voting. The proposed model outperformed the existing CNN models used for classification and attained an accuracy of 99.4% for the publicly available MIAS dataset. The major contributions of this research are: 1) Investigate two different feature fusion approaches (Fusion-1 & Fusion-2) for effective feature extraction, 2) Investigate the effect of ensemble learning using three different classifiers in the last block of CNN, 3) Embedding separable convolution and spatial dropout layers in mammographic classification network to study their performance impact, 4) Proposal of an end-to-end CAD system for effective classification of mammograms, which gives higher performance than existing techniques on the public datasets.

The remaining research paper is organized as follows: Section 2 details the proposed methodology, Section 3 shows experiments and results achieved, and section 4 concludes the research.

2. Methodology

Fig. 1 shows the general flow diagram of the proposed CAD system for breast cancer. The main process for tumor identification involves preprocessing, ROI extraction, features extraction, and classification, all discussed below.

2.1. Pre-processing

Pre-processing the raw mammogram image is important so that no ROI is missed in upcoming processes. For this purpose, an image enhancement algorithm is adopted to improve the quality of the mammogram image. In a raw mammogram, image edges are really important as they are the ones that will be used to extract the ROIs. For this reason, the unsharp masking technique is adopted. The unsharp masking technique enhances the edges of the image; therefore, a mask is created to blur the image. The blur kernel is defined as,

$$Blur = \frac{1}{2\pi s^2} \exp\left(-\frac{(X-m)^2 + (Y-m)^2}{2s^2}\right) \quad (1)$$

The blur kernel is convoluted with the input image, resulting in a blurred mammogram. This blurred image is subtracted from the

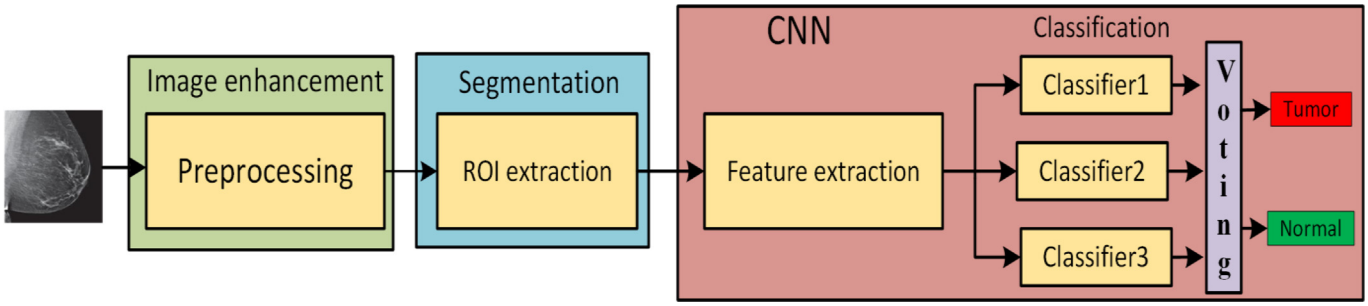

Fig. 1. General Flow Diagram of CAD System.

original mammogram image, which generates the mask. The mask is then added to the original image to achieve the edge-enhanced image. This can be numerically represented as,

$$v = y + q(x - y) \tag{2}$$

where x is the input mammogram image and y is the output of the low filter applied on a mammogram image. At the same time, the 'q' is the gain that should be real and positive as it acts as a scaling factor. The process is represented in the flow diagram shown in Fig. 2.

2.2. ROI extraction

We have an edge-enhanced image as an output from the pre-processing module. The ROI extraction algorithm proposed in this research depends on the edges of the mammogram image. The edges of the mammogram image are detected using a canny edge detector. A Gaussian filter is first applied to the enhanced image for the Canny edge detector, and the intensity gradient is found. The equation used of the gaussian filter is as follows,

$$G_{p,q} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(p-(l+1)^2 + (q-(l+1)^2}{2\sigma^2}\right) \tag{3}$$

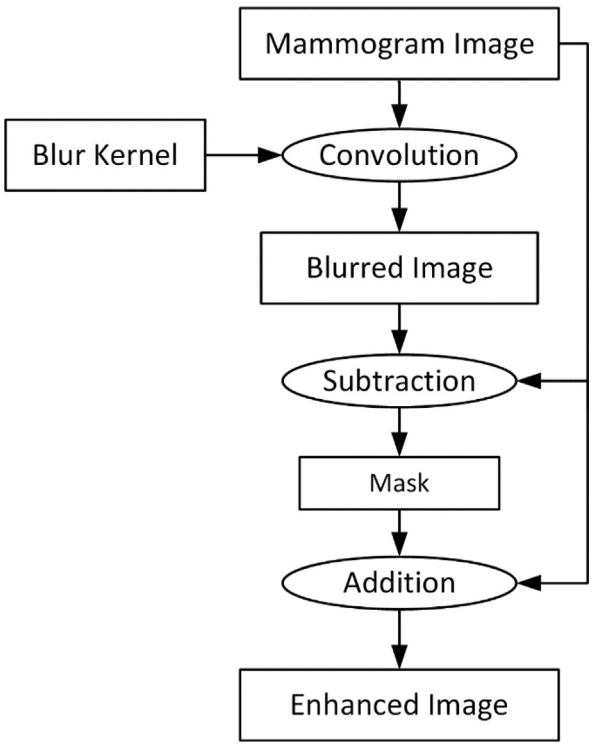

Fig. 2. Pre-processing algorithm.

where the value of p and q lies between 1 and (2 l + 1).

The following expressions calculate the intensity gradient magnitude and angle,

$$G = \sqrt{G_p^2 + G_q^2}$$

$$\theta = atan2(G_p, G_q)$$

After this process, non-maximum suppression is applied for thinning the edges of the image. After which double threshold is applied to get the potential and connected edges. Finally, edge tracking by hysteresis for the final edges of the image was performed. From these edges, the connected edges are separated and treated as a potential ROI. The samples of extracted normal and tumor ROIs can be seen in Fig. 3.

2.3. Proposed CNN architecture

The proposed CNN architecture plays the dual role of feature extraction as well as classification. The proposed CNN consists of four major blocks, three inside red-colored rectangle A and one inside rectangle B in Fig. 6. Each block in A consists of convolution, max-pooling, and dropout layers for feature extraction, while B has a flattening layer connected with three different blocks (sigmoid, SVM, RF). A combination of A and individual blocks in B will be considered a sub-network. This way, we have three sub-networks, and the last block of each sub-network is different. The first sub-network consists of two fully connected layers and a sigmoid layer for classification. The second sub-network consists of a support vector machine (SVM), while the third sub-network contains a random forest (RF) as classifiers.

The first block of all sub-networks consists of four layers where two layers are convolution with 64 filters of size 7 × 7, followed by a max-pooling layer and a spatial dropout layer. The second block consists of four layers again, but the depth-wise convolution layer is introduced this time. The third block consists of five layers where one depth-wise convolution layer is used with two simple convolution layers, a max-pooling layer and a spatial dropout layer. The last block consists of a flattening and classification layer where each sub-network is composed of a different classifier.

We investigated two different feature fusion approaches. Fig. 6 shows the details of Fusion-1 architecture. The max pool layer of each block contains particular 'details' of deep data. Local patterns are captured by the shallow layer, whereas global patterns are captured by the deep layers. Fusing all three layers (L1–L3) will enhance the classification model's discrimination ability by enriching the defining features of the input data. The particular strategy is to perform global average pooling (GAP) for each block and then concatenate the results in the GAP layer. GAP layer is further linked with the batch normalization (BN) layer, followed by the final classification block.

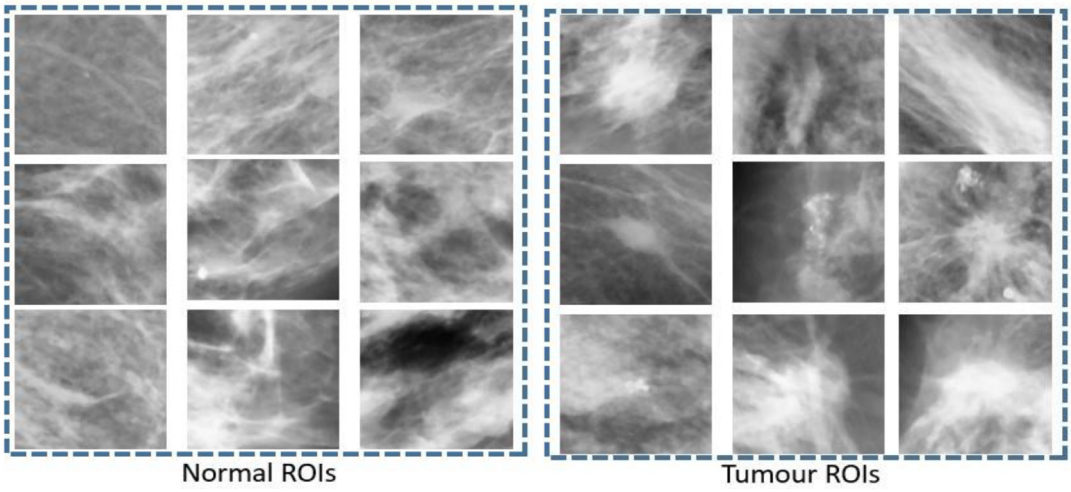

Fig. 3. Sample of Extracted ROIs.

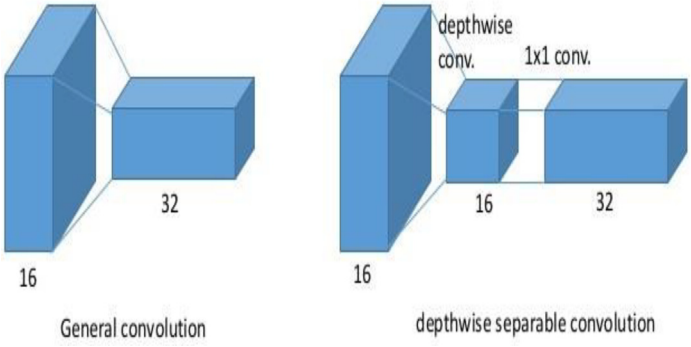

Fig. 4. Comparison between the general and depthwise convolution.

The data from the various channels will be strongly correlated, and the 1 × 1 convolution (Lin et al., 2013; Szegedy et al., 2015) will combine the cross-channel information. As a result, we develop the second deep fusion model (Fusion-2). The particular parameters setup of the BN and '1 × 1 conv' layers for L1-L3 branches are displayed in Fig. 7. The GAP details from each of the '1 × 1 conv' layers will be concatenated to create the longer GAP layer and forward the information to the BN layer. From here, the data is given to the last block as in Fusion-1.

Furthermore, the proposed architecture uses depth-wise convolution (Fig. 4) layers along with spatial dropout layers (Fig. 5), which have contributed significantly to the system performance. Using a depthwise convolution layer helps us have fewer parameters than the general convolution layer, resulting in a lesser chance of overfitting. Moreover, with the lower number of parameters, the weight assignment in backpropagation is done more effectively as unwanted parameters would be bypassed. Further depthwise convolution is a faster and cheaper process when compared to general convolution.

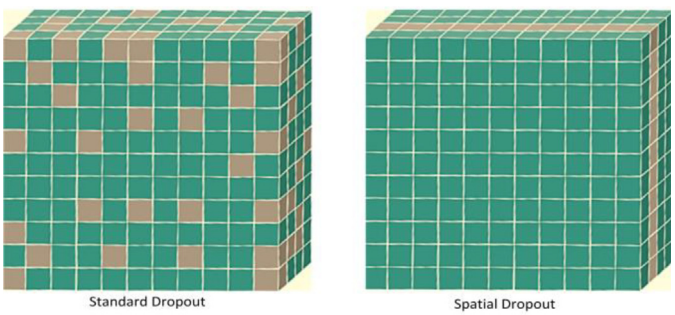

Fig. 5. The visualization of the standard and the spatial dropout layer.

The spatial dropout layer improves the generalization performance in a better manner than the standard dropout layer (Rehman et al., 2021). The spatial dropout layer drops the complete feature map rather than the individual elements, allowing only optimal feature maps to be part of a final feature vector used for classification purposes.

To check the value added by the ensemble approach, we used all the three classifiers individually, making our three sub-networks and altogether for the final prediction. The voting is applied to the predictions attained from individual sub-networks to get the final prediction regarding the ROI under consideration for the final model. The classifiers are discussed below.

2.3.1. Dense layer classifier

Two dense layers of size 512 and a sigmoid layer were used in the first subnetwork. The loss function used is categorical cross-entropy. The equation for the cross-entropy loss is as follows,

$$CE = \sum_{j}^{L} t_j log(f(p)_j) \quad (4)$$

where p is the input to the SoftMax layer and $f(p)_j$ is defined as,

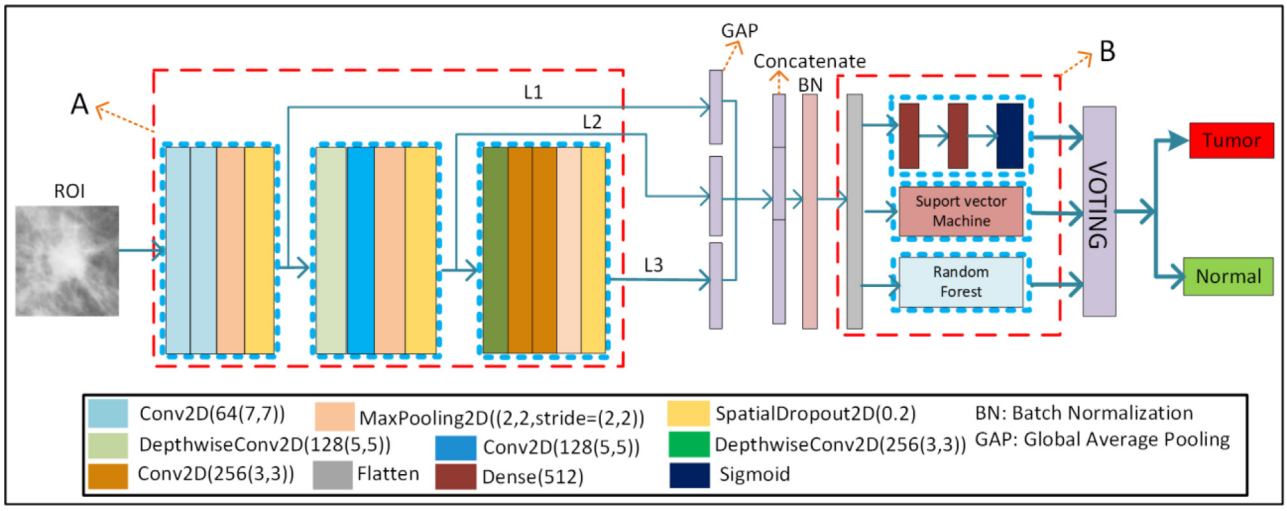

Fig. 6. In Fusion-1output of every block is passed through a GAP layer. The deep features are then concatenated into a larger GAP layer preceded by the BN, Flatten layer, and three Output classifiers to vote for the prediction.

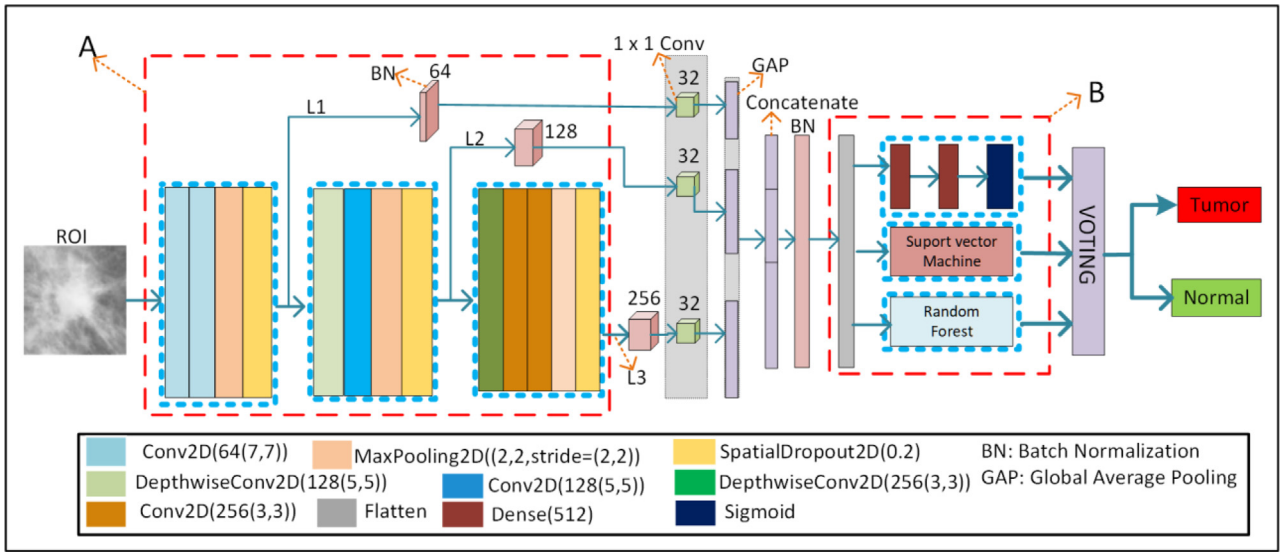

Fig. 7. Proposed Fusion-2 model. The output of every block is passed through BN, 1x 1 Conv, and GAP layers. The deep features are then concatenated into a larger GAP layer preceded by the BN, Flatten layer, and three Output classifiers to vote for the prediction.

$$f(p)_j = \frac{\exp(p_j)}{\sum_j^L \exp(p_j)} \tag{5}$$

2.3.2. SVM

SVM is a monitored ML technique that solves categorization difficulties. Every nth feature from the feature space is plotted as a point in n-dimensional space, with the value of each feature being the value of a certain coordinate in the SVM algorithm. Then it accomplishes classification by locating the hyper-plane that clearly distinguishes the two classes.

Consider 'u_i' as a collection of training samples and 'z_i' as their respective masks. Where $u_i \in R^d$, with'd' being the input image's dimension. Each image is assigned the class label $z_i \in \{1, 1\}$. The goal here is to create a hyperplane that splits the provided instances such that pixels from a particular class appear on a single face of the hyperplane. This hyperplane's equation is given below:

$$z_i\left(h\hat{A} \cdot u_i + c\right) > 0, i = 1, 2, 3, \tag{6}$$

Here c and h denote the offsets and weights to be calculated to satisfy Eq. (6). The data is considered linearly separable if there is a hyperplane for it. We can change'h' and'c' to make the distance between the hyperplane and the nearest data point 1/||h||. The optimization is carried out using Eq. (7).

$$z_i\left(h\hat{A} \cdot u_i + c\right) \geq 1, \tag{7}$$

Among all the hyperplanes that may be possible for the above equation, the best separating hyperplane is that plane farthest from the closest data location. The restriction of a margin of 1/||h|| has been imposed here, implying we must find a minimum $||h||^2$ for Eq. (7).

2.3.3. Random forest (RF)

RF utilizes the ensemble learning approach for regression and classification by integrating the impacts of several decision trees' outputs. The RF uses a voting approach to classify the testing data; the class with the most votes is regarded as the training data class.

For decision tree training, the RF uses the bootstrap aggregating or bagging approach. RF is a set of randomized classifying trees built using training data chosen at random. Random feature subsets for different tree nodes are picked, and the best parameters are chosen from them. The RF develops the decision tree utilizing the following technique.

1. Assume that RF is given N-dimensional data for training. First, RF will randomly choose samples with replacement; this information will be utilized to make bootstrap samples, used to form a decision tree.
2. M denotes the size of the original feature vector, and K∈(1, M) is a static parameter. For each node, the best configuration is selected from a random collection of its feature set.
3. This practice allows the tree to reach its full extent without pruning.

2.3.4. Majority voting

Following the classifiers' individual classifications, the most accurate outcome was obtained by applying majority voting to all three classification outputs. The following equation depicts the deployment of majority voting.

$$\text{Output} = Round\left(\sum_{k=1}^{3}\frac{f_k}{3}\right)$$

k denotes different classifiers (k = 1, 2, 3 denote dense classifier, SVM, and RF respectively). f is the classifier's decision. The value of f can be either 1 or −1, with 1 indicating normal ROI and −1 indicating tumor ROI. After computing f for all three classifiers, the ultimate conclusion was decided by a majority vote. If the Output > 0, the ROI is considered normal, and if Output < 0, the ROI is considered a tumor.

3. Experiments and results

3.1. Database

Mammography provides an excellent opportunity to identify the abnormality in the breast masses (Johns and Yaffe, 1987). In the medical field, there are two main categories of mammogram images called Screen Film Mammogram (SFM) and Full Field Digital Mammogram (FFDM) (Mustra et al., 2016). This study used a publicly available dataset called Mammogram Image Analysis Society (MIAS) (Suckling et al., 2015) of FFDM images. It is one of the most widely used datasets for research purposes. It provides metadata regarding mammography, which is made available after its annotation from experts in the field. There are 322 digital images in the MIAS dataset of 1024 × 1024. The number of normal and abnormal images is 208 and 114, respectively. As the proposed model is based on the deep neural network, data augmentation is used to increase the dataset size. The training split was augmented (by randomly flipping horizontally and vertically, rotation, varying scale, gamma correction) to increase the training dataset size to 4500. The second dataset used in our study is the BREAST CANCER DIGITAL REPOSITORY (BCDR) dataset. BCDR-D02 contains 230 biopsy-proven lesions of 162 patients, while BCDR-DN01 contains 200 normal mammograms of 48 patients. To validate our model on the BCDR dataset, we select 70 mammogram images containing 35 normal and 35 abnormal images.

3.2. Metrics

The proposed work uses three metrics to evaluate our model: sensitivity, specificity, and accuracy. These metrics, which range from 0 to 1, measure the performance of classification from various perspectives. The core values required to calculate these metrics are True Negatives (TN), False Negatives (FN), True Positives (TP), and False Positives (FP). The ability to determine a tumor ROI as tumor ROI will be known as sensitivity. It is the probability or chance that a tumor ROI would be detected as a tumor ROI. Below is the equation used for calculating the sensitivity,

$$Sensitivity = \text{TP}/(\text{TP} + \text{FN})$$

Specificity measures the probability or chance of how firmly a normal ROI image would be detected as the normal ROI. The term that is associated with the detection of normal ROI is called specificity.

$$Specificity = \text{TN}/(\text{TN} + \text{FP})$$

The accuracy of the technique depends on both sensitivity and specificity. The formulae for which is given below:

$$Accuracy = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

3.3. Model training

As discussed earlier, the MIAS dataset is used to train the model, which contains 322 images. The data augmentation was applied to the 120 × 120 ROIs after splitting the data into training, validation, and testing sets. 70% of the data is kept for training purposes, while 10% is kept as a validation set. The remaining 20% of the data was used for testing. After splitting the data into different sets, data augmentation is applied to increase the size of the training dataset. The data augmentation is applied later so that the trained model can be tested on a completely unseen dataset. Furthermore, we also evaluated our model on 70 (35 + 35) mammogram images from the BCDR dataset after applying preprocessing and ROI extraction.

To choose the most optimal parameters of the network, hyperparameter tuning is applied. The model uses Adam optimizer with a learning rate of 0.0001, β1 of 0.9, and β2 of 0.99, and a batch size of 24 for our final experiment. Early stopping is used based on validation loss. Early stopping helps to tackle the problem of overfitting, and it successfully gives the most optimal number of epochs for the given model. Moreover, all convolution and dense layers employ the exponential linear unit (eLu) activation function. The condition of the eLu activation function is as follows,

$$f(x) = \begin{cases} x, if x \geq 0 \\ \alpha(exp(x) - 1, otherwise \end{cases}$$

3.4. Ablation study

To understand the contribution made by the proposed Fusion-1, Fusion-2, ensemble learning, depthwise separable convolution, and spatial dropout layers, we carried out different experiments.

Table 1 demonstrates that the overall accuracy rate increases as the number of fusion branches grow. Theoretically, integrating additional information will improve classification results. Fusion-1(1–3) scores well across all assessment measures. The 1x1 conv layer will reveal the deep cross-channel information. The L1 will explore the local cross-channel features, while the L2 and L3 will focus on more 'complex' cross-channel features. The Fusion-2 got the high sensitivity and accuracy values than Fusion-1; this can well authenticate the cross-channel features and thus increase the classification results. Therefore for further experiments, we will consider our Fusion-2 approach. The ROC curve and AUC of both fusion models are shown in Fig. 8.

Next, we quantitatively compared the use of depthwise separable convolution with simple convolution and a spatial dropout

Table 1
Performance comparison of two models on MIAS dataset using ensemble learning approach for classification.

Model	Fused branches	Sensitivity	Specificity	Accuracy
Fusion-1	2,3	0.881	0.898	0.877
	1,2,3	0.907	0.864	0.929
Fusion-2	2,3	0.981	0.985	0.984
	1,2,3	**0.995**	**0.994**	**0.994**

layer with a simple dropout layer. We tried several combinations to have a strong comparison. Table 2 shows the details of different combinations. The performance with a combination of depth-wise convolution and spatial dropout has the best performance. To get the most optimized results and good performance from the CAD system, the authors have proposed to use the spatial dropout layer and depth-wise convolution layer in the proposed model. Depth-wise convolution has fewer parameters, resulting in a lesser chance of overfitting. Moreover, the weight assignment in backpropagation is done more effectively, and unwanted parameters would be bypassed. The spatial dropout layer improves the generalization performance, and the whole feature map is dropped if it doesn't contribute towards improving the system performance.

3.5. Comparison with other methods

The extracted ROIs were used for classification purposes. We used transfer learning and trained the VGG (Simonyan and Zisserman, 2014) and AlexNet (Krizhevsky et al., 2017) on the MIAS dataset for comparison. After the proper training of both state-of-the-art (SOTA) approaches, the ROIs were classified using VGGNet and AlexNet as well. Table 3 shows the classification results using three sub-networks of the proposed model individually, the proposed ensemble learning approach, VGGNet, and AlexNet.

As shown in Table 3, the proposed ensemble learning approach has better results than sub-networks 1, 2, and 3 alone. It is worth noting that our three subnetworks got better results than AlexNet and VGGNet on both datasets, which reveals the efficiency of feature fusion, depth-wise convolution and spatial dropout in the convolution layers for the final classification.

The classification process in breast cancer detection plays a vital role due to the high similarity index between the tumor and normal ROIs. The proposed model is eligible to distinguish the normal and tumor ROIs effectively and has attained an accuracy of 0.994, a specificity of 0.994, and a sensitivity of 0.995, using the MIAS dataset, while accuracy, specificity, and sensitivity of 0.985, 0.984, and 0.988 respectively on the BCDR test set. Moreover, it is important to state that the VGGNet has almost two times higher learnable parameters (138 M) than the proposed model (69 M), making VGGNet a slower evaluator for this particular research problem. The ROC curves of the suggested technique compared with VGGNet and AlexNet are shown in Fig. 9. It is evident from the graph that our suggested classification approach has outperformed other techniques. The results show that the proposed method can solve the small dataset problem and a high similarity index between the classes.

The proposed model got better performance compared with VGG and AlexNet. This may reveal that when the model goes deeper, the overfitting issues will occur, and the generalization ability for mammogram classification task may decline. Next, the local patterns may play a more important role in this particular classification task. This may explain why the deeper models VGG and AlexNet did not perform well in this task.

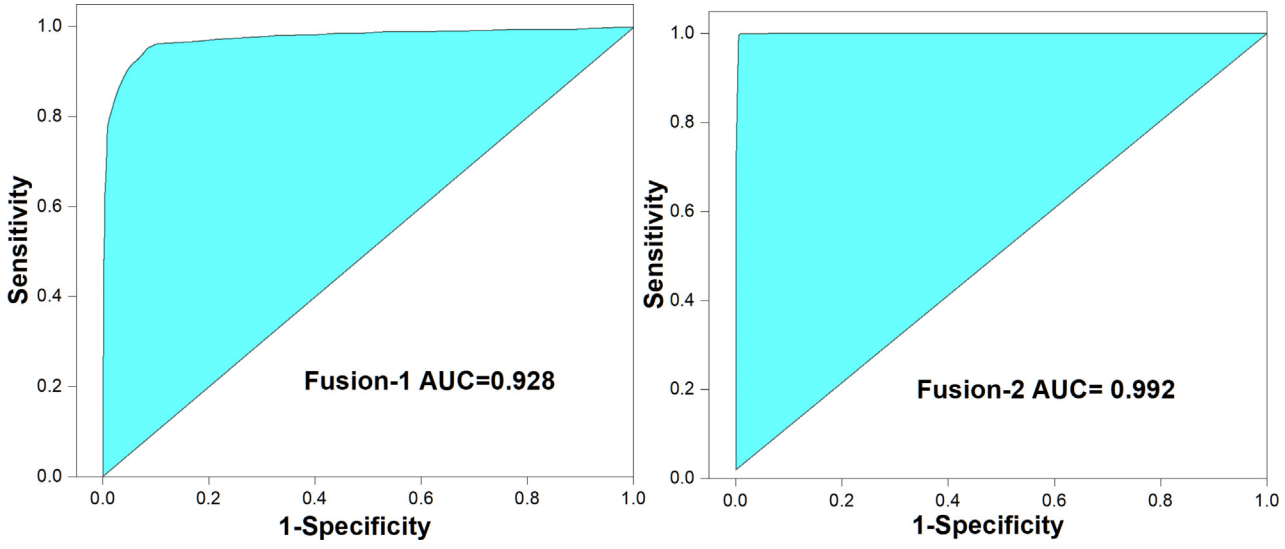

Fig. 8. Roc curve and AUC results of Fusion-1(1–3) and Fusion-2(1–3) from Table 1.

Table 2
Results of different combinations of simple dropout, spatial droput, simple and depthwise convolution for Fusion-2 model.

Combination	MIAS			BCDR		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
Conv + dropout	0.936	0.948	0.942	0.905	0.901	0.895
Dethwise-Conv + dropout	0.972	0.978	0.975	0.922	0.915	0.919
Dethwise-Conv + Spatial dropout	**0.995**	**0.994**	**0.994**	**0.988**	**0.984**	**0.985**
Conv + Spatial dropout	0.957	0.951	0.954	0.938	0.942	0.944

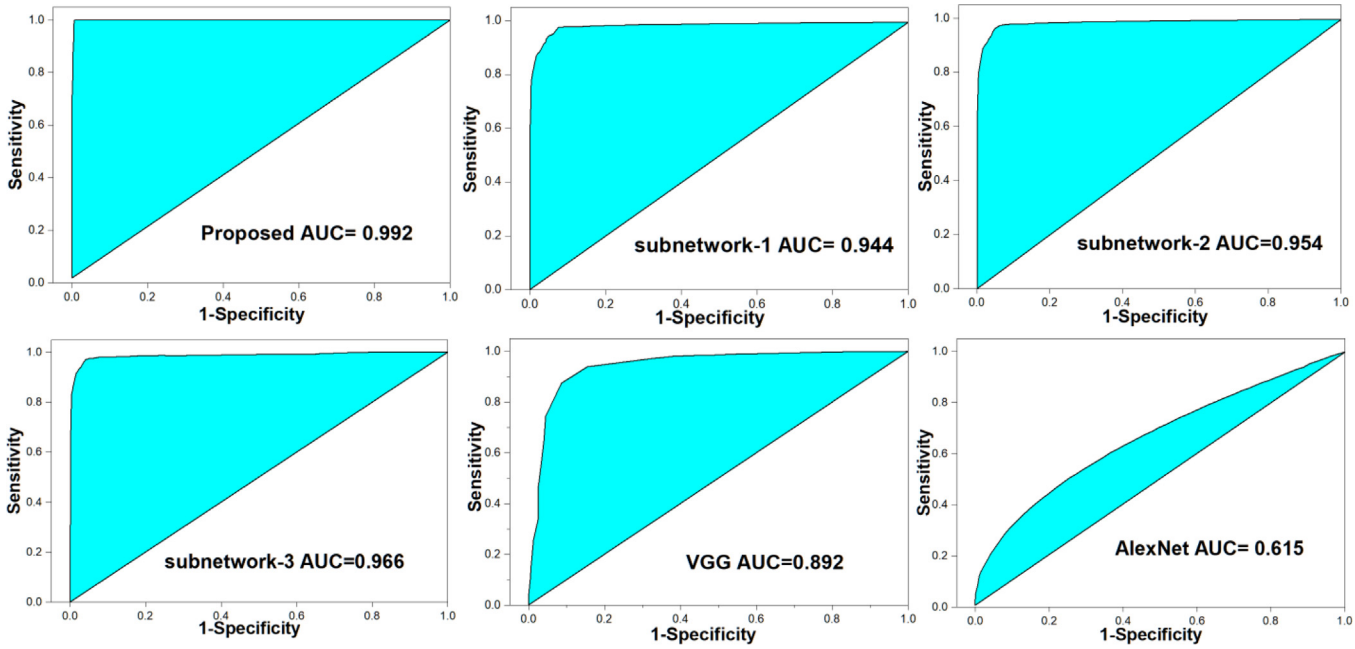

Fig. 9. Comparative Analysis of the proposed model, 3 subnetworks,VGG and AlexNet using AUC and ROC curves.

Table 3
Performance comparison of the proposed model with SOTA VGGNet and AlexNet.

Model	MIAS			BCDR		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
AlexNet	0.624	0.754	0.721	0.604	0.739	0.715
VGGNet	0.827	0.862	0.892	0.806	0.814	0.851
Sub-Network 1	0.973	0.981	0.976	0.968	0.975	0.974
Sub-Network 2	0.957	0.964	0.961	0.949	0.958	0.955
Sub-Network 3	0.976	0.968	0.973	0.969	0.961	0.966
Proposed	**0.995**	**0.994**	**0.994**	**0.988**	**0.984**	**0.985**

Table 4
Performance comparison of the proposed model with existing techniques.

Reference	CAD Scheme	Sensitivity	Specificity	Accuracy
Mabrouk et al. (2019)	SVM	0.650	0.900	0.750
Mabrouk et al. (2019)	KNN	0.650	0.800	0.770
Mabrouk et al. (2019)	ANN (Automatic)	0.980	0.940	0.960
Duraisamy and Emperumal (2017)	DL-CNN + FRCN	0.987	1.00	0.981
Rampun et al. (2018)	LQP + SVM	0.767	0.857	0.869
Bajaj et al. (2019)	BEMD features	1.00	0.929	0.950
Proposed		0.995	0.994	0.994

Table 4 shows the performance comparison of the proposed model with existing techniques present in the literature. As shown in Table 4, the previously proposed techniques have a high difference in specificity and sensitivity, which indicates that the models are biased towards one class. However, the proposed model has a minute difference, which means the proposed model treats both classes equally. The highest sensitivity is achieved by bi-dimensional empirical mode decomposition (BEMD) features. Still, unfortunately, this scheme failed to achieve high specificity, which resulted in lower accuracy than the proposed model. The Local Quinary Pattern (LQP) features used with SVM resulted in a high difference in sensitivity and specificity.

4. Conclusion and future work

Breast cancer can be cured if detected at an early stage. The conventional system for diagnosing this deadly disease is time-intensive and sensitive to unintentional errors. This paper proposes an end-to-end CAD system for the classification of breast cancer in mammogram images, consisting of image preprocessing, ROI extraction, and classification steps. The most important part of the CAD model is the feature extraction and classification, for which the proposed model used feature fusion and ensemble learning-based DCNN. Feature fusion is used to extract useful features from the ROIs, and ensemble learning is used to classify the ROIs for final prediction. Moreover, the proposed model uses spatial dropout, depth-wise convolution layers, and data augmentation to overcome the issues of overfitting and data scarcity. The proposed model attained an accuracy of 0.994, a specificity of 0.994, a sensitivity of 0.995 on the MIAS dataset, and accuracy, specificity, and sensitivity of 0.985, 0.984, and 0.988, respectively on the BCDR test set.

In the future, we want to investigate the deep fusion learning for identifying the contributions of distinct branches. We presume

that merging specific layers of patterns with different weights might improve model robustness.

Declarations

Conflict of interest

The authors would like to confirm that they have no conflicts of interest, financial, or others.

Data availability

In our experiments, the Mammogram Image Analysis Society (MIAS) dataset and BREAST CANCER DIGITAL REPOSITORY (BCDR) dataset were used to assess the efficiency of the suggested approach. The MIAS dataset is publicly available at *https://peipa.essex.ac.uk/pix/mias/all-mias.tar.gz*, while the BCDR dataset can be freely downloaded from *https://bcdr.eu/information/downloads* for research purposes.

References

Agarwal, R. et al., 2019. Automatic mass detection in mammograms using deep convolutional neural networks. J. Med. Imaging 6 (3), 031409.

Arevalo, J. et al., 2016. Representation learning for mammography mass lesion classification with convolutional neural networks. Comput. Methods Programs Biomed. 127, 248–257.

Bajaj, V. et al., 2019. Computer-aided diagnosis of breast cancer using bi-dimensional empirical mode decomposition. Neural Comput. Appl. 31 (8), 3307–3315.

Cai, H. et al., 2019. Breast microcalcification diagnosis using deep convolutional neural network from digital mammograms. Comput. Math. Methods Med. 2019.

Cai, X. et al., 2021. Breast cancer diagnosis by convolutional neural network and advanced thermal exchange optimization algorithm. Comput. Math. Methods Med. 2021.

D'Angelo, G., Palmieri, F., 2020. Knowledge elicitation based on genetic programming for non destructive testing of critical aerospace systems. Future Generation Comput. Syst. 102, 633–642.

D'Angelo, G., Palmieri, F., 2021. Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction. J. Network Comput. Appl. 173, 102890.

Duraisamy, S., Emperumal, S., 2017. Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier. IET Comput. Vision 11 (8), 656–662.

Guo, G., Razmjooy, N., 2019. A new interval differential equation for edge detection and determining breast cancer regions in mammography images. Systems Sci. Control Eng. 7 (1), 346–356.

Hudson, J., A Partially Observable Markov Decision Process for Breast Cancer Screening. 2019.

Huynh, B.Q., Li, H., Giger, M.L., 2016. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J. Med. Imaging 3 (3).

Jalalian, A. et al., 2013. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clin. Imaging 37 (3), 420–426.

Jiao, Z. et al., 2016. A deep feature based framework for breast masses classification. Neurocomputing 197, 221–231.

Johns, P.C., Yaffe, M.J., 1987. X-ray characterisation of normal and neoplastic breast tissues. Phys. Med. Biol. 32 (6), 675–695.

Khan, S. et al., 2016. Optimized Gabor features for mass classification in mammography. Appl. Soft Comput. 44, 267–280.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90.

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

Lin, M., Q. Chen, S. Yan, Network in network. arXiv preprint arXiv:1312.4400, 2013.

Liu, Q. et al., 2020. Computer-aided breast cancer diagnosis based on image segmentation and interval analysis. Automatika 61 (3), 496–506.

Mabrouk, M.S., Afify, H.M., Marzouk, S.Y., 2019. Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques. Ain Shams Eng. J. 10 (3), 517–527.

Mann, R.M. et al., 2008. Breast MRI: guidelines from the European society of breast imaging. Eur. Radiol. 18 (7), 1307–1318.

Moayedi, F. et al., 2010. Contourlet-based mammography mass classification using the SVM family. Comput. Biol. Med. 40 (4), 373–383.

Mustra, M., Grgic, M., Rangayyan, R.M., 2016. Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. Med. Biol. Eng. Compu. 54 (7), 1003–1024.

Oeffinger, K.C. et al., 2015. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 314 (15), 1599–1614.

Oliver, A. et al., 2012. Automatic microcalcification and cluster detection for digital and digitised mammograms. Knowl.-Based Syst. 28, 68–75.

Olsen, O., Gøtzsche, P.C., 2001. Cochrane review on screening for breast cancer with mammography. Lancet 358 (9290), 1340–1342.

Rampun, A. et al., 2018. Breast density classification using local quinary patterns with various neighbourhood topologies. J. Imaging 4 (1), 14.

Razmjooy, N., Estrela, V.V., Loschi, H.J., 2020. Entropy-based breast cancer detection in digital mammograms using world cup optimization algorithm. Int. J. Swarm Intel. Res. (IJSIR) 11 (3), 1–18.

Rehman, M.U. et al., 2021. m6A-NeuralTool: convolution neural tool for RNA N6-Methyladenosine site identification in different species. IEEE Access 9, 17779–17786.

Rehman, M.U. et al., 2021. BrainSeg-Net: brain tumor MR image segmentation via enhanced encoder–decoder network. Diagnostics 11 (2), 169.

Rehman, M.U., Chong, K.T., 2020. DNA6mA-MINT: DNA-6mA modification identification neural tool. Genes 11 (8), 898.

Shi, P. et al., 2019. Deep learning from small dataset for BI-RADS density classification of mammography images. 2019 10th International Conference on Information Technology in Medicine and Education (ITME). IEEE.

Siegel, R.L. et al., 2022. Cancer statistics, 2022. CA A Cancer J Clinicians.

Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

Smart, C.R. et al., 1995. Benefit of mammography screening in women ages 40 to 49 years. Current evidence from randomized controlled trials. Cancer 75 (7), 1619–1626.

Suckling, J. et al., 2015. Mammographic image analysis society (mias) database v1, 21.

Szegedy, C. et al., 2015. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Tang, J. et al., 2009. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. IEEE Trans. Inf Technol. Biomed. 13 (2), 236–251.

ur Rehman, M. et al., 2018. Classification of skin lesion by interference of segmentation and convolotion neural network. 2018 2nd International Conference on Engineering Innovation (ICEI). IEEE.

Wang, J. et al., 2016. Discrimination of breast cancer with microcalcifications on mammography by deep learning. Sci. Rep. 6, 1–9.

Wei, L. et al., 2005. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. IEEE Trans. Med. Imaging 24 (3), 371–380.

Wei, J. et al., 2005. Computer-aided detection of breast masses on full field digital mammograms. Med. Phys. 32 (9), 2827–2838.

Whang, J.S. et al., 2013. The causes of medical malpractice suits against radiologists in the United States. Radiology 266 (2), 548–554.

Zhu, W. et al., 2017. Deep multi-instance networks with sparse label assignment for whole mammogram classification. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer.