



# Detection of abnormalities in mammograms using deep features

Nasrin Tavakoli<sup>1</sup> · Maryam Karimi<sup>2</sup> · Alireza Norouzi<sup>1</sup> · Nader Karimi<sup>1</sup> · Shadrokh Samavi<sup>1</sup> · S. M. Reza Soroushmehr<sup>3</sup>

Received: 31 July 2019 / Accepted: 6 December 2019 / Published online: 14 December 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

The mortality rate of breast cancer can be reduced by early diagnosis and treatment. The computer-aided diagnosis (CAD) systems can effectively help physicians identify the early stage of breast cancer. The primary tool for such systems is mammograms. Usually, these images lack high quality. Furthermore, due to the irregular shape of masses, their size variability, and the apparent similarity of the masses and other dense regions of the breast tissue, it is difficult to detect and diagnose the masses. Although many image processing techniques have been presented for diagnosis of breast masses, they have not been quite successful, and this problem has been retained as a challenge yet. In this paper, a method for classifying breast tissues into normal and abnormal (i.e., cancerous) is proposed, which is based on a deep learning approach. It mainly contains a new convolutional neural network (CNN) and a decision mechanism. After a preprocessing phase, a block around each pixel is fed into a trained CNN to determine whether the pixel belongs to normal or abnormal tissues. This results in a binary map for the input suspicious tissue. Afterward, as a decision mechanism, a thresholding technique is applied to a central block on the produced binary map to label it. The new architecture of the CNN, the training scheme of the network, employing the CNN for classifying the pixels of the suspicious regions rather than the entire input, exerting an effective decision mechanism on the output of the CNN and the learning of the model parameters, helped us achieve superior results compared to state-of-the-art methods by reaching 95 and 94.68 percent for AUC and accuracy, respectively.

**Keywords** Deep learning · Convolutional neural network · Breast tissue classification · Computer-aided diagnosis · Mammogram

## 1 Introduction

Breast cancer is a fatal disease that originates in breast tissue. It could affect women more often, though men may also be affected (Wajid and Hussain 2015). According to the American Cancer Society (2011), almost one out of every nine women is estimated to develop breast cancer. Early diagnosis and treatment can significantly reduce the mortality rate (Buciu and Gacsadi 2011). However, interpretation of mammograms, as a primary tool for the breast cancer diagnosis, is challenging because of the subtle nature of mammographic abnormalities, poor quality of mammograms, the rarity of expert radiologists, and boredom for construing large numbers of images in limited time (Chakraborty et al. 2012). The computer-aided diagnostic systems can help the specialists greatly to recognize the early stages of breast cancer. Mammography has been shown as the most effective and reliable tool to diagnose breast cancer at an initial stage (Davies and Dance 1990; Lau and Bischof 1991; Siddiqui

---

✉ S. M. Reza Soroushmehr  
ssoroush@umich.edu

Nasrin Tavakoli  
ntavakoli88@gmail.com

Maryam Karimi  
m.karimi@sci.sku.ac.ir

Alireza Norouzi  
nrzi.alireza@gmail.com

Nader Karimi  
nader.karimi@cc.iut.ac.ir

Shadrokh Samavi  
samavi96@cc.iut.ac.ir

<sup>1</sup> Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran

<sup>2</sup> Shahrekord University, Shahrekord, Iran

<sup>3</sup> University of Michigan, Ann Arbor, USA

et al. 2005). On mammograms, dense breast tissue, such as breast masses or tumors, look white and hence in healthy women with dense breasts mammogram is not a reliable tool for cancer diagnosis (American College of Radiology 2013). The irregular shape of masses, their size variability, and complexity of the breast tissue make it difficult to separate the mass from other dense regions of the breast tissue. Also, sometimes the size of the lesion is too small to be seen by experts. Many methods have been proposed to deal with these challenges. However, the accuracy of breast mass detection still needs improvement. Some samples of normal and abnormal breast tissues are shown in Fig. 1. A brief overview of different existing methods is provided in Sect. 2.

In this paper, we propose a deep feature-based scheme that principally contains a convolutional neural network (CNN) and a decision mechanism to classify breast tissues into normal and abnormal. We have a preprocessing phase which eliminates irrelevant information from the image and enhances the contrast of the mammogram. Then, a new architecture for a block-based CNN is presented. This network is trained on a large number of normal and abnormal blocks from the images of the training set. In the test phase, the suspicious regions of the images, we call each of them a region of interest (ROI), are extracted and fed into the CNN. Afterward, the CNN classifies the pixels of each ROI into normal and abnormal with labels '0' and '1' to generate a binary map. In the next stage, an efficient decision mechanism based on a thresholding technique is applied to the central block of the resulted binary map to label the inputted ROI. The appropriate size of the block and the threshold parameter are experimentally determined using the training data, and the best values are applied in the test phase. Unlike many methods that employ rescaling the image for feeding to CNN, the ROIs in the proposed method are not rescaled to preserve the quality of the image. Furthermore, many existing approaches exploit CNN to classify the ROIs, whereas we employ a new CNN to classify the pixels of the ROIs first. Afterward, classified ROIs are assigned to another stage by applying an effective decision mechanism

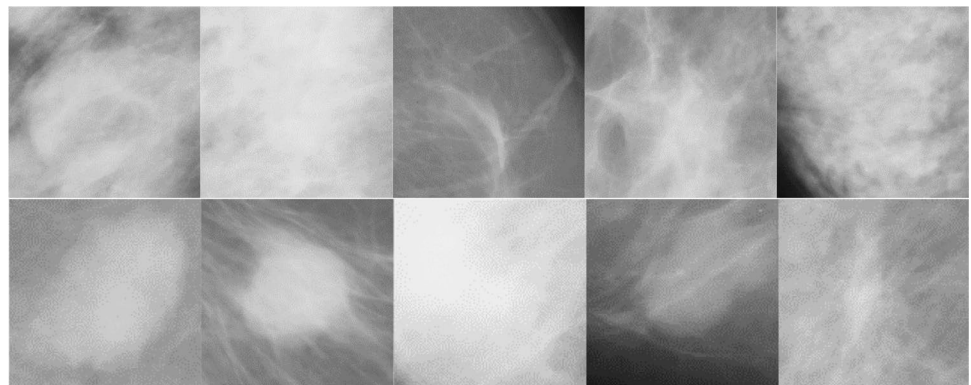
on the output of the CNN. The obtained results show the superiority of the proposed algorithm compared to state-of-the-art methods.

The rest of this paper is organized as follows. A brief overview of the existing techniques is supplied in Sect. 2. The different stages of our proposed method, including preprocessing, processing data by CNN, and a decision mechanism are described in Sect. 3. After introducing the database, in Sect. 4, the proposed method is compared with other mammogram classification methods. Furthermore, the influential model parameters are varied in the proposed framework, and their effects are examined in the results. Besides, the effect of preprocessing, investigating a CNN-based procedure instead of the thresholding-based way, the environment of simulations and computation time are presented in this section. Finally, Sect. 5 concludes the paper with a direction for future research.

## 2 Literature review

In the literature, several types of research have been introduced for the detection and diagnosis of masses in mammograms. In a bunch of studies, low-level or medium-level features such as margin or the shape of masses are extracted (Cristianini et al. 2002) first. Then, the features are presented into different kinds of classifiers to categorize masses. Moayedi et al. (2010) employ the contourlet transform to gain its coefficients as features and performed a classification based on the Support Vector Machine (SVM) family. In Buciu and Gacsadi (2011), directional features are extracted after filtering the images by Gabor wavelets and eventually, Proximal Support Vector Machines (PSVM) are used to classify them. Agrawal et al. (2014) use saliency-based segmentation, namely GBVS (Harel et al. 2007) to extract the suspicious regions in mammograms. Then, a large number of features are extracted from the segmented areas, and subsequently, 154 features are selected to be classified by an SVM classifier (Chang and Lin 2011). Feature extraction

**Fig. 1** Instances of breast tissues. The top row represents samples for the normal case, while the bottom row illustrates samples for the abnormal breast tissues



based on curvelet transform and moment theory is exerted in Dehahbi et al. (2015) to describe the images. Next, a K-Nearest Neighbor (K-NN) classifier is used to distinguish between normal and abnormal breast tissues. In our previous work (Tavakoli et al. 2017) a supervised discriminative dictionary learning approach is applied on DSIFT (Dense Scale Invariant Feature Transform) features. Meanwhile, a linear classifier is simultaneously learned with the dictionary to classify the sparse representations.

Moreover, Tosin et al. (2018) develop the curvelet transform to extract shape features from the ROIs while texture features are extracted using the Local Binary Pattern (LBP) algorithm. The K-NN algorithm also is employed to classify the extracted features. In Jen and Yu (2015), the breast image is segmented by a thresholding technique. After applying gray-level quantization on the segmented breast image, five first-order statistical intensities and gradients features are extracted from a suspicious ROI. Then, feature difference matrices are created from the extracted features, and principal component analysis (PCA) (Pearson 1901) is used to help the determination of feature weights. In the method presented by Khuzi et al. (2009), after a preprocessing stage, three segmentation methods namely local threshold (Cheng et al. 2006), k-mean (Cheng et al. 2006), and Otsu (Otsu 1979) are checked to extract intended ROIs. Afterward, textural features of ROIs are extracted by using gray level co-occurrence matrices (GLCM). Finally, a decision tree employs these features to detect masses. GLCM is also used for feature extraction in the method proposed by Ancy and Nair (2017) where an SVM model performs mammograms classification. A CAD system, based on extracting features using exact Gaussian–Hermite moments, is introduced in Eltoukhy et al. (2018).

The obtained feature vector is presented to K-NN, random forests, and AdaBoost classifiers to differentiate between normal and abnormal lesions. The authors in Chakraborty et al. (2018) present a multiresolution analysis of tissue pattern orientation to categorize masses. Although a wide range of traditional features seems to make good descriptions of an image, a considerable gap exists between these features and cognitive behaviors of physicians (Doi 2007).

Therefore, strategies should be based on how radiologists look at medical images. To judge whether a medical image is normal or abnormal, physicians combine different levels of knowledge with previous experience in similar tasks. However, it seems complicated to identify the hierarchical sense of mass images and information processing of the human brain by traditional features and related methods (Jiao et al. 2015). Deep learning (Nielsen 2015) is a machine learning paradigm that tries to mimic the human brain by transferring semantic information from lower levels to higher levels. Deep learning has been playing a significant role in the

academic society and has caused an immense change in the field of big data and artificial intelligence (Jiao et al. 2015). A type of deep architecture that is particularly applicable in the field of image processing is the CNN (Nielsen 2015) that consists of two primary layers, which are the convolutional layer and the pooling layer. The convolutional layer calculates the output of the neurons that are connected to the local area at the input by sharing weights and biases and the pooling layer subsamples the output of the convolutional layer and decreases the dimensionality of the data (Ertosun and Rubin 2015). CNN can automatically learn suitable image features for different applications more efficiently compared to hand-crafted features utilized by traditional machine learning approaches (Ertosun and Rubin 2015).

Many methods use CNN to detect abnormal areas in mammographic images (Jiao et al. 2015; Ertosun and Rubin 2015; Jaffar 2017; Abbas 2016; Bay et al. 2006; Guo et al. 2010; Schmidhuber 2015; Qiu et al. 2016; Lo et al. 2002; Zhang et al. 2018). Jiao et al. (2015) extract two groups of deep features from two different layers called high-level and middle-level features. Then, intensity information and extracted deep features are combined by a decision mechanism. After that, the outcomes of classifiers based on different features are jointly analyzed to characterize the types of test images. Ertosun and Rubin (2015) propose a system with two modules called classification engine and localization engine. A deep CNN classifies mammograms to containing a mass or not. Then, a regional probabilistic approach based on a deep learning network localizes the mass within the image. In Jaffar (2017), images are first resized, and then a CNN is used for extraction of features, and finally, an SVM classifies the features. Another deep learning-based approach in Abbas (2016), extracts two descriptors; speed-up robust features (SURF) (Bay et al. 2006) and local binary pattern variance (LBPV) (Guo et al. 2010) from each mass. These descriptors are transformed into deep invariant features (DIFs) (Schmidhuber 2015) in a supervised and unsupervised manner through a multilayer deep-learning architecture. A fine-tuning step completes the determination of the features, and the final decision is made via a soft-max linear classifier. The method in Qiu et al. (2016), utilizes an eight-layer deep learning network for automatic feature extraction and a multiple layer perceptron (MLP) classifier for feature categorization. The MLP classifier generates a classification score to predict the likelihood of an ROI depicting a malignant mass. Lo et al. (2002) designed a multiple circular path convolution neural network (MCPCNN) for the analysis of tumor and tumor-like structures. In this way, each suspected tumor area is divided into sectors, and the defined mass features for each sector is computed independently. These sector features are used on the input layer and coordinated by convolution kernels of different sizes. In Zhang et al. (2018), a nine-layer CNN is proposed. In this

method, three activation functions, moreover, six pooling techniques are compared and eventually, the results show the combination of a parametric rectified linear unit (ReLU) and a rank-based stochastic pooling performs the best.

Most methods utilize CNNs for feature extraction and then perform classification by traditional machine learning approaches such as SVM (Jiao et al. 2015; Jaffar 2017), MLP (Qiu et al. 2016), Neural Network (NN) (Guan and Loew 2017), and so on. In this paper, CNN is employed for classifying all pixels of a suspicious region to obtain a binary map first. Then, the resulted binary maps are labeled by a threshold-based decision mechanism. Moreover, we apply a CNN-based approach rather than the threshold-based technique for the classification task and analyze the outcomes of the two paths. Unlike some methods (Ertosun and Rubin 2015), which use well known pre-designed CNNs, we performed a variety of experiments to find the appropriate CNN architecture for our work. Moreover, many existing schemes (Ertosun and Rubin 2015; Jaffar 2017; Abbas 2016; Qiu et al. 2016; Zhang et al. 2018; Guan and Loew 2017) down sample or resize the input image to reduce the computational complexity of the CNN, which decreases the quality of mammograms. To overcome this problem, we consider a window around each pixel of the breast tissue as the CNN input. Therefore, not only the image quality is preserved, but also a large number of input blocks for training the CNN are generated. We will show that the preprocessing stage, including contrast enhancement and pectoral muscle suppression, significantly improves the performance, in contrast to some methods (Ertosun and Rubin 2015; Abbas 2016; Guan and Loew 2017) that are lacking this step.

### 3 Proposed method

The block diagram of our proposed method is shown in Fig. 2. The proposed method consists of three main steps: (a) preprocessing, (b) classifying the pixels of the inputted

suspicious region (ROI) by a CNN, and (c) Assigning a single label of ‘normal’ or ‘abnormal’ to each ROI by a decision mechanism.

The image preprocessing eliminates irrelevant areas from the image and enhances the contrast of the mammogram. In the second stage, CNN is trained by random blocks selected from normal and abnormal tissues of the training set. In the test phase, the trained CNN is employed to classify the suspicious regions (ROIs) pixels of the test images. Therefore, a binary map for each ROI is obtained. In the third stage, the ROIs based on their binary maps are classified into normal or abnormal breast tissues. In the following, we explain all the mentioned steps in detail.

#### 3.1 Pre-processing

Due to random block selection from the entire image in the network training phase, we need to delete irrelevant areas from the image to supply more accurate results. Based on the experiments we performed, preprocessing significantly increases the accuracy of the results. To eliminate the background area in mammogram images, such as high-intensity rectangular label, tape artifact, noise, etc. the preprocessing technique is essential (Narain Ponraj et al. 2011). Also, pectoral muscle in the upper corner of a breast has a similar intensity as dense structures in breast tissue such as abnormal masses, fibro-glandular disc (in a Medio Lateral Oblique (MLO) view of a mammogram) (Maitra et al. 2012). Moreover, mammograms are grayscale images with poor contrast. Using contrast enhancement techniques makes intensities of pixels have better distribution in the image histogram. Therefore, similar to our previous work (Tavakoli et al. 2017), the pre-processing stage includes four main steps: (1) breast region extraction, (2) pectoral muscle suppression, (3) mask creation, and (4) contrast enhancement. In the following, we explain each of the mentioned steps.

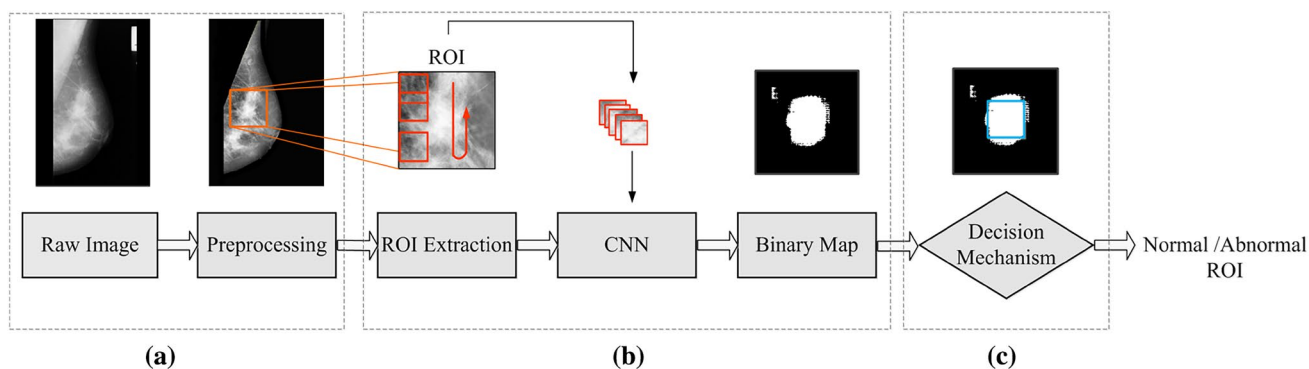


Fig. 2 The diagram of the proposed method

### 3.1.1 Breast region extraction

To identify the breast object, we employ Otsu’s thresholding method (Otsu 1979) to find the adaptive threshold  $t_o$  corresponding to each image. To find more precise boundaries of the breast region, we modify Otsu’s threshold by multiplying  $t_o$  by a constant value,  $0 < \alpha' < 1$ , and calculate  $t_f = \alpha' \times t_o$  as the final threshold to binarize mammograms. Then a flat, disk-shaped structuring element with a radius of two pixels is used to dilate the image. After that, the largest disjoint component in the binary image is selected as the breast region. This resulting mask (Fig. 3d) is then multiplied by the original image, and thus, the breast region is extracted (Fig. 3e). The effect of this process at each step is shown in Fig. 3.

### 3.1.2 Pectoral muscle suppression

We remove the pectoral muscle region from mammograms according to the method proposed in Jen and Yu (2015). To specify the location of the muscle, the orientation of the breast is first determined by a method explained in Jen and Yu (2015). In this method, at least four horizontal reference lines are considered at intervals of 1.4 or 1.8 image width depicted as dotted lines in Fig. 4a that pass across both sides of the breast contour at eight cross points  $((x_i, y_i)$  where  $1 \leq i \leq 8$ ). If the four points of the cross (i.e.  $(x_i, y_i)$  for  $5 \leq i \leq 8$ ) are placed on a vertical line, the pectoral muscle is on the same side as the four points. If there is no vertical permutation of cross points, we get the slope of the straight-line L called ‘s’ passing through two upper cross points on a non-vertical contour curve as seen in Fig. 4a. If the slope s is greater than zero, the breast orientation is specified as the right of the image; vice versa, if it is less than zero the pectoral muscle is assessed as left of the image.

To remove pectoral muscle, after finding the breast orientation, the contrast of the image is enhanced by using gamma correction equalization (Jen and Yu 2015). Next, a modified Otsu’s thresholding method is applied to the enhanced image to obtain a binary image. Then, dilation and erosion

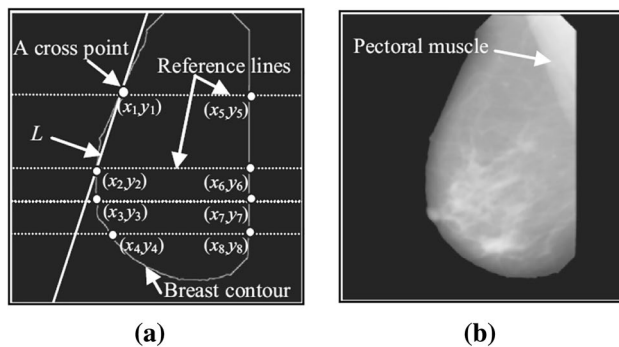


Fig. 4 Determining breast orientation, a a breast contour, b a breast image (Jen and Yu 2015)

operations are applied to the binary image, and finally, according to the breast orientation, the candidate component in the upper corner of the image is eliminated from the original image. Figure 5 displays the output of this step in a typical example.

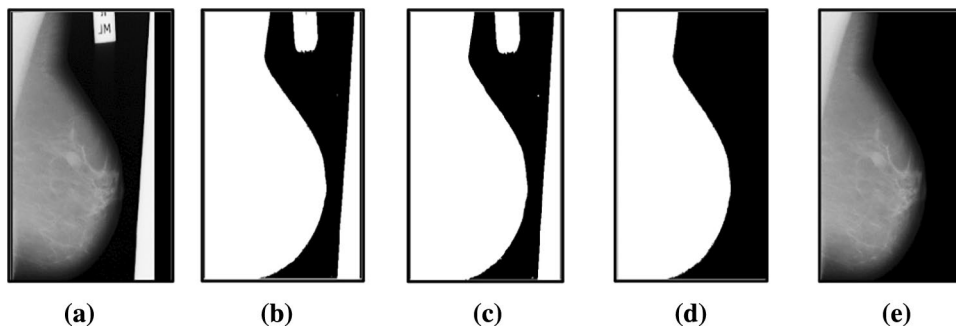
### 3.1.3 Mask creation

To train the CNN in the proposed method, it is necessary to determine the labels of normal and abnormal blocks as a ground-truth. Also, in the test set, this information is used to compare the predicted results with targets. In the database used for experiments, the abnormality in a mammogram is specified via both the center coordinates of the lesion and a radius of its expansion. In the proposed method, a mask is created for each image based on this knowledge. In this mask, the background pixels, the foreground pixels (i.e., pixels related to the healthy tissue of the breast) and the abnormal pixels are characterized by black, white, and gray colors respectively. The sample of the resulted mask can be seen in Fig. 7.

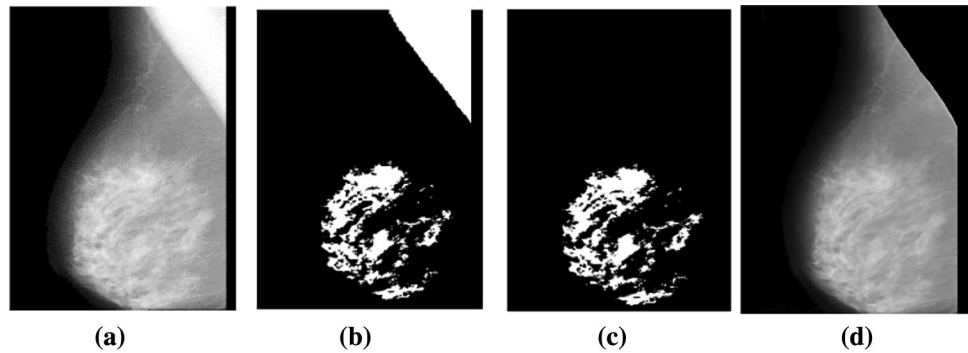
### 3.1.4 Contrast enhancement

We utilize the Contrast Limited Adaptive Histogram Equalization (CLAHE) (Pizer et al. 1987) algorithm to enhance

Fig. 3 Breast region extraction, a original image, b the binary image, c the image dilation, d binary breast region extraction, e extracted breast region



**Fig. 5** Pectoral muscle elimination, **a** enhanced image by gamma correlation, **b** the obtained binary image by using modified Otsu's thresholding, **c** removing the candidate component in the corner of the binary image, **d** Pectoral muscle elimination from the original image



the contrast of mammograms. This approach is a common technique for enhancing medical images (Wajid and Husain 2015). CLAHE divides an image into contextual blocks called tiles and then exploits histogram equalization (HE) (Pizer et al. 1987) to each tile. Afterward, it makes a histogram for each tile by using a specific number of bins and clips the histogram at a specified threshold. Then, it maps each region according to the new histogram results. This technique leads to artificial effects at tile boundaries. Therefore, the bilinear interpolation method combines neighboring tiles. The contrast, particularly in homogeneous areas, can be limited to prevent reinforcing any noise that might exist in the image. The effect of the contrast enhancement algorithm, CLAHE, on mammograms, is shown in Fig. 7.

### 3.2 Processing data by convolutional neural network

This stage aims to classify pixels of suspicious regions of the image (ROIs) by a block-based CNN. First, the preprocessed images are randomly divided into training and test sets. The CNN is trained on several normal and abnormal random blocks taken from the training mammograms. In the test stage, like most of the existing methods (Moayedi et al. 2010; Buciu and Gacsadi 2011; Dehahbi et al. 2015; Tosin et al. 2018; Eltoukhy et al. 2018; Guan and Loew 2017; Setiawan et al. 2015; Chougrad et al. 2018), the ROIs are extracted firstly from the test set. Afterward, To classify the ROI pixels, a block is considered around each pixel and fed to the trained CNN to determine the labels of central pixels. These labels form a binary map as the output of the CNN. In the next step, based on the binary map, it is decided whether an ROI is normal or abnormal. In the following, CNN architecture, network training, and test are explained in more detail.

#### 3.2.1 Architecture of CNN

The new architecture of the network proposed in this paper is shown in Fig. 6. Four convolution layers are used for experimentation with 32, 64, 128 and 256 filters and kernel sizes of  $7 \times 7$ ,  $5 \times 5$ ,  $3 \times 3$ , and  $3 \times 3$  respectively. The Rectified Linear Unit (ReLU) is used as the activation function. Moreover, Batch normalization (Ioffe and Szegedy 2015) approach which accelerates training network speed, is applied after the convolution layers. Both the weights and biases are initialized randomly by Glorot and Bengio's method (Glorot and Bengio 2010). Then, the stochastic gradient descent (SGD) (Bottou 2010) is employed to minimize the cross-entropy over the training set. The batch size for the SGD function is set to 64 with a momentum of 0.8. The last convolutional layer is followed by a max-pooling layer with a kernel size of 2. Afterward, a regularization technique, namely "dropout," is used in the fully-connected layer by setting the probability to 0.5. Overfitting can be reduced by using dropout to prevent complex co-adaptations on the neurons (Hinton et al. 2012). Then, a flatten layer exploited to create a single long feature vector for the fully-connected layer with 128 hidden neurons. In the end, another fully connected layer with one neuron and sigmoid function is applied for classification.

#### 3.2.2 Network training

To train the block-based CNN, normal and abnormal blocks are selected from the labeled mammograms and fed to the CNN. In mammography images, abnormal regions are minor parts of the image as opposed to normal regions. This means that by a purely random selection of training blocks, the majority of the selected blocks would belong to the normal tissues. This unbalanced distribution of training blocks would reduce the learning performance of the network. In the proposed method, blocks with  $b \times b$  pixels that are randomly selected from normal and abnormal tissues of the mammogram are used as CNN inputs for network training (Fig. 7). To avoid the problem of having unbalanced training data, in the selection of training samples for CNN, 50% of

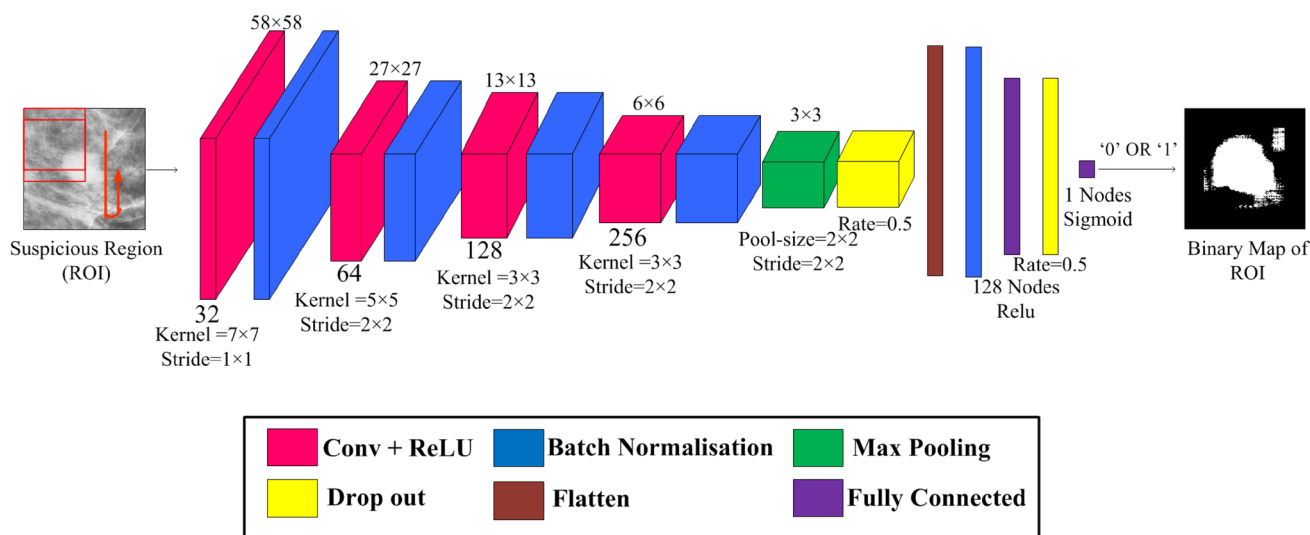


Fig. 6 The architecture of the proposed network

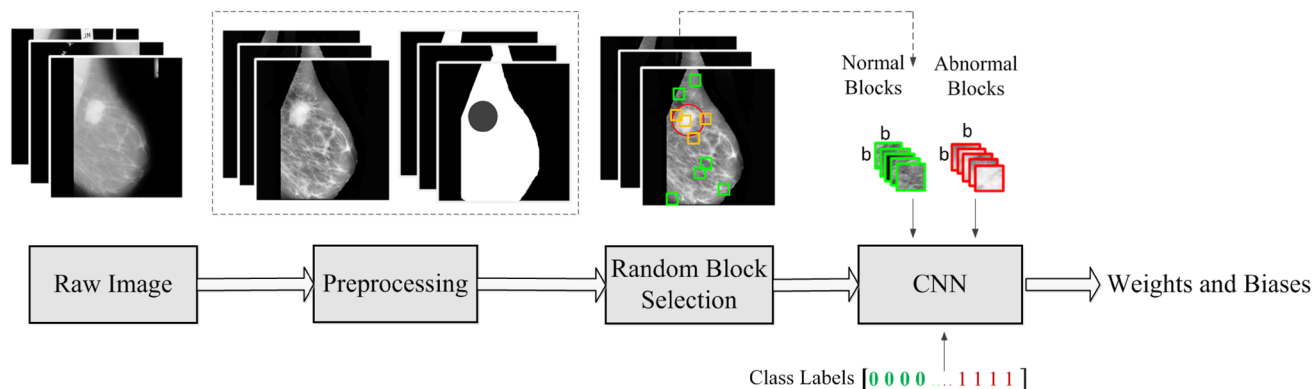


Fig. 7 The training of the network in the proposed scheme

blocks are randomly selected from normal tissues, and the other 50% are extracted from abnormal tissues. The selection is made based on the available ground truth of the images and generated masks of mammograms in Sect. 3.1.3. Also, ‘0’ and ‘1’ labels are assigned to the central pixels of the normal and abnormal blocks, respectively as targets.

### 3.2.3 Network testing

As mentioned earlier, suspicious regions as ROIs are selected in the test stage from the test set to examine the trained CNN. For this purpose,  $q \times q$  squares surrounding the centers of tumorous masses are taken as abnormal ROIs and the same size squares also randomly selected inside breast tissues as normal ROIs to contain all normal tissue types equally (fatty, fatty–glandular, and dense–glandular) (Moayedi et al. 2010; Buciu and Gacsadi 2011; Dehahbi et al. 2015; Tosin et al. 2018; Eltoukhy et al. 2018; Guan

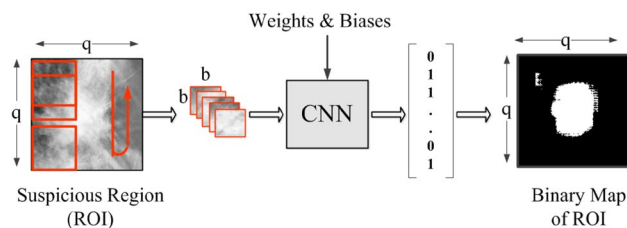


Fig. 8 The testing phase of the proposed system

and Loew 2017; Setiawan et al. 2015; Chougrad et al. 2018). Figure 1 illustrates five samples per class (case). To classify the pixels of testing ROIs, a block with the size of  $b \times b$  pixels around each pixel is considered and fed into the trained network. The learned block-based CNN assigns ‘0’ or ‘1’ label to each block center. Accordingly, a binary map with the same size of each ROI is resulted (Fig. 8). The

intensity values of zero and one in the binary map in Fig. 8 show the pixels that the CNN detects normal and abnormal respectively. This output of the CNN is post-processed by a decision mechanism introduced in the next section to determine if each ROI is normal or not.

### 3.3 Decision mechanism

To assign a single label to each ROI, a decision mechanism is performed here. For this purpose, we consider a block in the center of the obtained binary map as seen in Fig. 9c, d. Then ROI labeling is done based on a threshold value, called ' $\alpha$ .' If the number of abnormal pixels in the central block (with the size of  $h \times h$ ) exceeds the threshold of  $\alpha$ , that ROI will be labeled as abnormal. Otherwise, it will be labeled as healthy breast tissue.

## 4 Experimental results

In our experiments, 70% of the dataset, 145 normal and 70 abnormal mammograms, were used as the training data, and the remaining 30% including 64 normal and 30 abnormal mammograms, were used for testing the method. To overcome the shortage of breast images for training CNN, we trained our CNN on blocks. Hence, we elicit 450,000 normal and 450,000 abnormal  $64 \times 64$ -pixel blocks ( $b = 64$ ) from training mammogram set while 30,000 blocks from each class are randomly selected to train the CNN. To investigate the ROIs of the test set, the square areas with the size of  $128 \times 128$  pixels are extracted ( $q = 128$ ). The extracted ROI is fed into the CNN in the form of the blocks (with the size of  $64 \times 64$  pixels) around each pixel. Consequently, a binary map, with the size of the ROI ( $128 \times 128$  pixels), is resulted

as the output of the CNN. To decide on the central block size ( $h$ ) and the threshold parameter ( $\alpha$ ) in the decision mechanism stage, we performed several experiments on training data (Fig. 10) and chose  $h = 32$  and  $\alpha = 0.6$ . Finally, a label was assigned to each ROI as normal or abnormal.

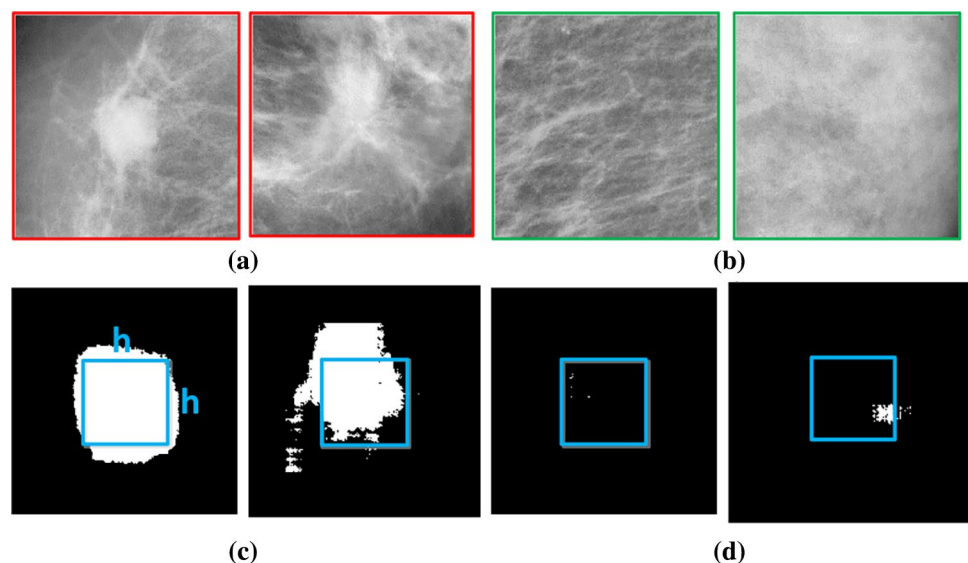
### 4.1 Dataset

The mammograms used in this experiment are taken from the mini mammography database of MIAS (Suckling et al. 1994) that is available online (2019). This database has 322 mammograms in Medio Lateral Oblique view that the actual size of all images is  $1024 \times 1024$  pixels and has been digitized with 50-micron pixel. Also, they are held as 8-bit gray-scale images with 256 different gray levels. The dataset includes 209 normal and 113 abnormal mammograms. In terms of types of breast tissue, images are classified into three groups of fatty, fatty-glandular, and dense-glandular. Also, Masses in abnormal cases are categorized into five classes: Calcification, Circumscribed masses, Speculated masses, Ill-defined masses, and Architectural distortion. Furthermore, the severity of abnormality can be malignant or benign. The locations of mass centers and their radius determine a ground truth for each image of the dataset.

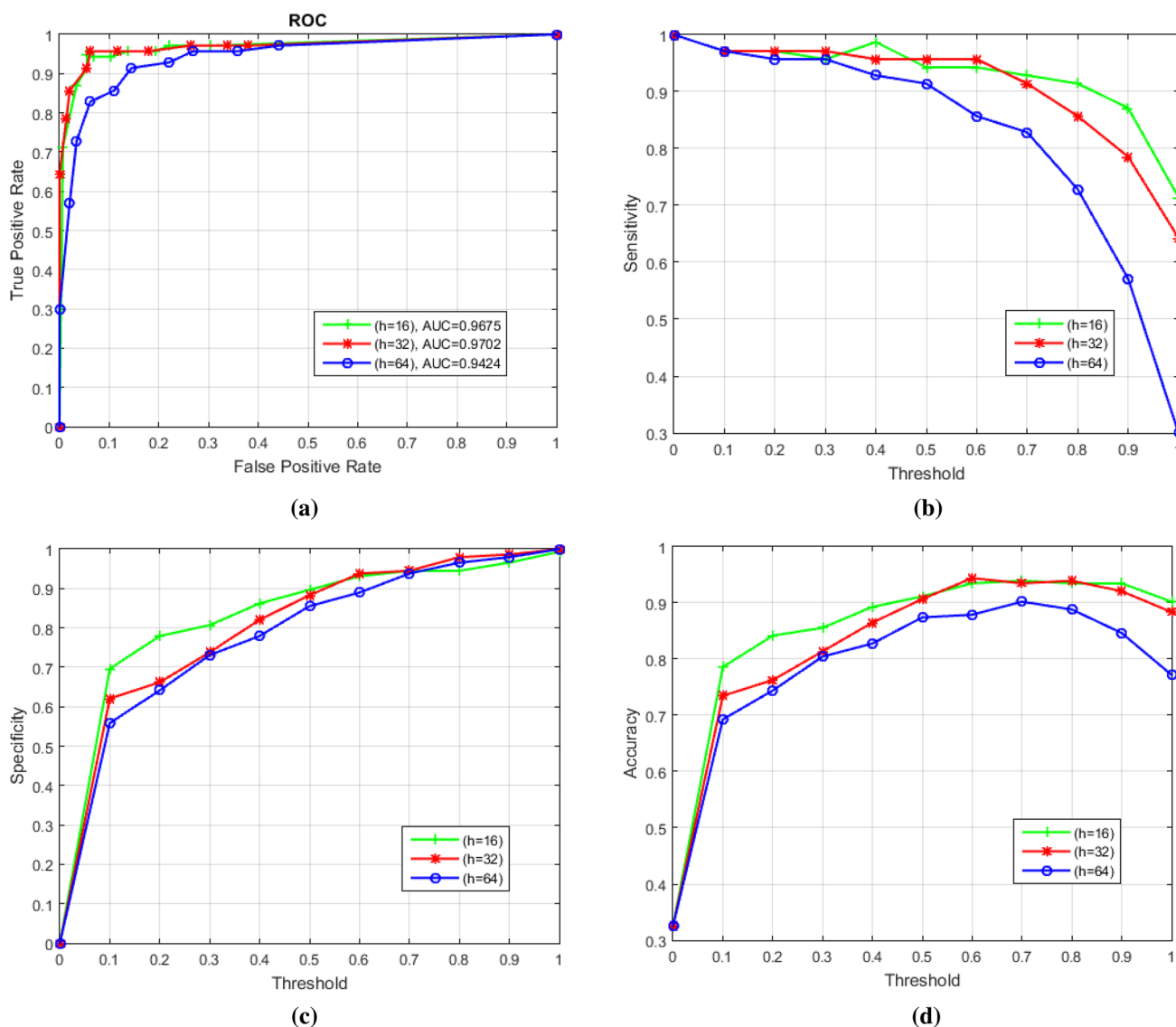
### 4.2 Evaluation methodology

Four different measures, including Accuracy, Sensitivity, Specificity, and Area Under Curve (AUC), have been used to evaluate the performance of the proposed framework. These standard evaluation criteria are represented in Table 1. In this table, True Positive (TP) defines the number of accurately classified ROIs, which are abnormal. True Negative (TN) specifies the number of accurately classified normal

**Fig. 9** Samples of input mammograms ROIs considering the central block on their resulted binary maps, **a** abnormal ROIs, **b** normal ROIs, **c** binary maps of the abnormal ROIs with the central block, **d** binary maps of normal ROIs with the central block







**Fig. 10** The curves of **a** ROC, **b** sensitivity, **c** specificity and **d** accuracy for different value of dimensions  $h$  of the central block on the training data

**Table 1** Standard evaluation criteria

Specificity	$\frac{TN}{TN + FP}$
Sensitivity	$\frac{TP}{TP + FN} = \frac{TP}{T_s}$
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$

ROIs. The other two counterparts, False Positive (FP) and False Negative (FN) respectively represent the number of inaccurately classified ROIs, which are normal or abnormal. Also, Trues ( $T_s$ ) and Falses ( $F_s$ ) are the numbers of all abnormal and normal ROIs used in the testing stage.

AUC is the most usual measure to assess overall discrimination. It is a number between zero and one that shows the area under the Receiver Operating Curve (ROC). An AUC

value of 0.5 indicates a random prediction (poor discrimination), and a value of 1 is ideal for a predictor (excellent discrimination). In the figure of the ROC curve, the vertical and horizontal axes are True Positive Rate (TP-rate) and False Positive Rate (FP-rate). TP-rate (sensitivity) measures the proportion of actual positives that are correctly identified, and FP-rate is the proportion of all negatives that still yield positive test outcomes. A larger area under this curve stands for better classification performance.

### 4.3 Effects of model parameters

In this section, the parameters of the decision mechanism stage are varied on the training data to assess their effects on

the final results. To better illustrate the process of changing the indicators by variations of the parameters, the graph of each of the metrics is drawn in Fig. 10. We set the parameters in such a way that the best results were achieved for the diagnosis task on the training data. Initially, the suspicious regions (ROIs) of the training set are elicited in the same manner explained in Sect. 3.2.3 to feed into the CNN and their binary maps are generated. Parameter  $h$  that shows the dimensions of the central block of a binary map (Sect. 3.3), is evaluated in three different sizes of 16, 32, and 64 pixels. The AUC values attain 0.9675, 0.9702 and 0.9424 for sizes 16, 32 and 64 pixels of  $h$  respectively (Fig. 10a). Also, the threshold parameter  $\alpha$ , that the ROI labeling is applied based on it, is changed between 0 and 1. If the number of abnormal pixels in the central block of the binary map exceeds the threshold of  $\alpha$ , the ROI will be recognized as abnormal breast tissue; otherwise, it will be normal. Figure 10 shows when  $\alpha$  increases step by step, both of the specificity, and accuracy, usually increase too. But the sensitivity criterion remains constant or decreases. Furthermore, a larger area under the ROC curve (close to 1) stands for better classification performance. Therefore, after performing different experiments and according to the obtained results of Table 2, we set  $\alpha=0.6$  and  $h=32$  for the proposed system to gain our best results.

#### 4.4 Investigating a CNN-based approach instead of the thresholding-based technique in decision mechanism stage

In this section, the power of the diagnosis system is surveyed when CNN is exploited rather than the proposed decision mechanism stage. For this purpose, a second CNN including four convolution layers with 32, 64, 128 and 256 filters is proposed where the size of each kernel is  $3 \times 3$ , and their activation function is the ReLU. Furthermore, the Batch Normalization procedure (Ioffe and Szegedy 2015) is employed after the filters in convolution layers. The last convolutional layer is followed by a max-pooling layer with a kernel size of 2. After that, the dropout regularization technique (Hinton

et al. 2012) is exerted by setting the probability to 0.3. Then, a flatten layer is utilized to form a single long feature vector for the next fully-connected layer with 128 hidden neurons. In the end, we use the dropout manner again by assigning 0.5 to the probability. Afterward, another fully connected layer with one neuron and sigmoid function is applied for classification. The adaptive learning rate method (Adadelata) (Zeiler 2012) is employed as the optimizer, which adapts learning rates based on a moving window of gradient updates. The training process of this CNN is continued for 100 epochs and each epoch runs in 21 s. In the test stage, it takes 11.20 ms to label each ROI binary map into normal or abnormal.

The obtained results for the classification of the ROIs in this way are compared with the thresholding-based method. As seen in Table 3, the achieved results in the two methods are slightly different. As the sensitivity metric focuses on the existence of cancerous regions, it is a preferable index for performance evaluation of cancer detection methods. In other words, higher values mean more sensitivity of the approach to the existence of cancerous masses in medical applications. Therefore, we choose the thresholding-based path in the decision mechanism due to its higher sensitivity index.

#### 4.5 Effect of preprocessing

In this final experiment, we validate the effectiveness of the preprocessing stage including the contrast enhancement and the elimination of pectoral muscle. To do so, we performed the same experiments once without contrast enhancement and another time without pectoral muscle suppression.

**Table 3** Comparison of two methods for the decision mechanism

Decision mechanism	Sensitivity	Specificity	Accuracy	AUC
Threshold-based	<b>93.33</b>	95.31	94.68	<b>0.95</b>
CNN-based	90.00	<b>96.87</b>	94.68	0.93

The bold values indicate the best result for each metric

**Table 2** Investigating the effects of model parameters; threshold parameter and dimensions of the central block on training data

' $\alpha$ '	h = 16			h = 32			h = 64		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
0.6	94.28	93.10	93.49	<b>95.71</b>	<b>93.79</b>	<b>94.42</b>	85.71	88.96	87.90
0.7	92.86	94.48	93.95	91.42	94.48	93.49	82.85	93.79	90.23
0.8	91.42	94.48	93.49	85.71	97.93	93.95	72.85	96.55	88.84
0.9	87.14	96.55	93.49	78.57	98.62	92.09	57.14	97.93	84.65
1	71.42	99.31	90.23	64.28	1.00	88.37	0.30	1.00	77.21

The bold values indicate the best results obtained on the training data

**Table 4** Evaluating the impact of pre-processing on the outcomes of the proposed method

Methods	Sensitivity	Specificity	Accuracy	AUC
Proposed method without contrast enhancement	80.00	84.37	82.97	0.82
Proposed method without pectoral muscle suppression	86.66	89.06	88.29	0.88
Proposed method with contrast enhancement and pectoral muscle suppression	93.33	95.31	94.68	0.95

The obtained results for each experiment are listed in Table 4. As seen, improving the image contrast has a greater impact on the efficiency of the method.

### 4.6 Computation time

Simulations in this study were carried out in Spyder integrated development environment (IDE) in Python 3.6. All experiments were performed on a cluster node with 4.00 GHz Intel Core i7 CPU and 32 GB RAM in windows 7.

The training phase is performed in offline mode. The stage of the CNN training is terminated after 150 epochs and the epoch takes 820 s (about 13.5 min) to run. In the testing phase, the computation time to evaluate each ROI by the trained CNN is 67 s. Besides, the decision mechanism based on thresholding takes 17.41 ms to label each binary map.

### 4.7 Performance comparison

As already mentioned, comparing the results of methods is performed based on four standard evaluation criteria in Table 2. The sensitivity criteria refer to the ability of the method to correctly identify those patients with breast cancer. The specificity metric refers to the ability of the test to identify the patients without the disease, and Accuracy correctly shows the capability to correctly recognize the checked patients (both with and without breast cancer). The

performance of the proposed framework is compared at different aspects with the other state-of-the-art methods in Table 5, in which all, detection of abnormality is performed on the extracted ROIs in the form of the squares from the MIAS dataset. Our proposed method, with the presentation of a new architecture for a block-based CNN, took advantage of deep features to classify the pixels of breast tissues. Also, it then employed an effective decision mechanism on the outputs of CNN to gain acceptable outcomes in the diagnosis of abnormality. Therefore, from the comparison in Table 5, our approach has better efficiency in important criteria, accuracy, and AUC by achieving 94.68 percent and 0.95, respectively. In other metrics, specificity and sensitivity, we also reach satisfactory results that way we are the second-best in specificity metric by 95.31 percent, and we have about 4 percent difference in sensitivity criteria in the best way by 93.33 percent. These outcomes show the good ability of the system in detecting abnormal areas among current frameworks.

## 5 Conclusion

In this paper, we proposed a new scheme based on deep features for classifying breast tissues as normal or abnormal. The proposed method consisted of three main parts, including preprocessing, a CNN, and a decision mechanism. First,

**Table 5** Comparison of performance among different mass detection schemes in terms of sensitivity, specificity, accuracy, and area under curve (AUC) on MIAS dataset

Method	Extracted features	Sensitivity	Specificity	Accuracy	AUC
Khuzi et al. (2009)	GLCM (Otsu segmentation)	70.00	100.00	–	0.84
Moayedi et al. (2010)	Feature extraction based contourlet transform	–	–	85.90	–
Buciu and Gacsadi (2011)	Directional features using Gabor wavelets	97.56	60.86	84.37	0.79
Eltoukhy et al. (2018)	Feature extraction using exact Gaussian-Hermite moments	–	–	90.56	0.88
Tosin et al. (2018)	Feature extraction based curvelet transform and LBP algorithm	96.00	93.70	94.20	–
Dehahbi et al. (2015)	Feature extraction based curvelet transform and moment theory	–	–	91.27	–
Lo et al. (2002)	Feature extraction by a multiple circular path convolution neural network (MCPCNN)	–	–	–	0.89
Zhang et al. (2018)	Feature extraction by convolutional neural network with a parametric rectified linear unit (ReLU) and a rank-based stochastic pooling	93.40	94.60	94.00	–
Proposed	Feature extraction by convolutional neural network	93.33	95.31	<b>94.68</b>	<b>0.95</b>

The bold values show the superiority of the proposed method compared to others

a preprocessing stage was employed to prepare mammograms for later stages. Then, a block-based CNN with new architecture was trained on the randomly selected normal and abnormal blocks from training images. After the training stage, the CNN could classify the pixels of each ROI into zero (for normal pixel) and one (for abnormal pixel). So that a binary map with the same size of the ROI results. Next, in the decision mechanism step, a central block was considered on the produced binary map that if the number of abnormal pixels in this central block exceeded a threshold, the ROI was labeled abnormal and vice versa. Several experiments on the training data found the suitable size of the central block on the binary map and the threshold parameter of this step, and the best values were adjusted by analyzing different curves. The new architecture of the CNN, the training way of the CNN, applying an efficient decision mechanism and the learning of the model parameters, led to good results in our work. The capability of this framework was investigated on the MIAS database with 209 normal and 113 abnormal mammograms so that the accuracy and Area Under Curve criteria in our method were achieved to 94.68% and 0.95 respectively. In our future research, we will attempt to find a better alteration of CNN to help attain more descriptive features and design a decision mechanism for the reduction of false positive in the framework. Also, we would like to evaluate our scheme on other clinical databases for the stability test of abnormality detection.

## References

- Abbas Q (2016) DeepCAD: a computer-aided diagnosis system for mammographic masses using deep invariant features. *Computers*. <https://doi.org/10.3390/computers5040028>
- Agrawal P, Vatsa M, Singh R (2014) Saliency-based mass detection from screening mammograms. *Signal Processing* 99:29–47
- American Cancer Society (2011) Breast cancer facts and figures 2011–2012. American Cancer Society Inc, Atlanta
- American College of Radiology (2013) BI-RADS ATLAS Mammography. Reporting System. <https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/Mammography-Reporting.pdf> Accessed 13 June 2019.
- Ancy CA, Nair LS (2017) An efficient CAD for detection of tumor in mammograms using SVM. In: Proceedings of ICCSP 2017 International Conference on Communication and Signal Processing, Chennai, India.
- Bay H, Tuytelaars T, Van Gool L (2006) Surf: Speeded up robust features. In: Proceedings of the 9th European Conference on Computer Vision, 3951, pp. 404–417, Graz, Austria.
- Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT' 2010, Physica-Verlag HD, pp 177–186
- Buciu I, Gacsadi A (2011) Directional features for automatic tumor classification of mammogram images. *Biomed Signal Process Control* 6(4):370–378
- Chakraborty J, Rangayyan RM, Banik S, Mukhopadhyay S, Desautels JEL (2012) Statistical measures of orientation of texture for the detection of architectural distortion in prior mammograms of interval cancer. *J Electron Imaging* 21(3):1–13
- Chakraborty J, Midya A, Rabidas R (2018) Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. *Expert Syst Appl* 99:168–179
- Chang CC, Lin CJ (2011) LIB SVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):1–27
- Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN (2006) Approaches for automated detection and classification of masses in mammogram. *J Pattern Recogn Soc* 39(4):646–668
- Chougrad H, Zouaki H, Alheyane O (2018) Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs Biomed* 157:19–30
- Cristianini N, Shawe-Taylor J, Lodhi H (2002) Latent semantic kernels. *J Intell Inform Syst* 18(2–3):127–152
- Davies DH, Dance DR (1990) Automatic computer detection of clustered calcifications in digital mammograms. *Phys Med Biol* 35(8):1111–1118
- Dehabhi S, Barhoumi W, Zagrouba E (2015) Breast cancer diagnosis in digitized mammograms using curvelet moments. *Comput Biol Med* 64:79–90
- Doi K (2007) Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 31(4):198–211
- Eltoukhy MM, Elhoseny M, Hosny KM, Singh AK (2018) Computer aided detection of mammographic mass using exact Gaussian-Hermite moments. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-018-0905-1>
- Ertoşun MG, Rubin DL (2015) Probabilistic Visual Search for Masses Within Mammography Images using Deep Learning. In: Proceedings of BIBM 2015 IEEE International Conference on Bioinformatics and Biomedicine
- Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feed forward neural networks. *Proc Int Conf Artif Intell Stat* 9:249–256
- Guan S, Loew M (2017) Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks. *IEEE Applied Imagery Pattern Recognition Workshop (AIPR), USA*. <https://doi.org/10.1109/aipr.2017.8457948>
- Guo Z, Zhang L, Zhang D (2010) Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recogn* 43(3):706–719
- Harel J, Koch C, Perona P (2007) Graph-based visual saliency. *Adv Neural Inf Process Syst* 19:545–552
- Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving neural networks by preventing co-adaptation of feature detectors. [arXiv:1207.0580](https://arxiv.org/abs/1207.0580)
- Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:448–456
- Jaffar AM (2017) Deep Learning based computer aided diagnosis system for breast mammograms. *Int J Adv Comput Sci Appl (IJACSA)* 8(7):286–290
- Jen Ch, Yu Sh (2015) Automatic detection of abnormal mammograms in mammographic images. *Expert Syst Appl* 42(6):3048–3055
- Jiao Z, Gao X, Wang Y, Li J (2015) A deep feature based framework for breast masses classification. *Neurocomputing* 197:221–231
- Khuzi AM, Besar R, Wan Zaki WMD, Ahmad NN (2009) Identification of masses in digital mammogram using gray level co-occurrence matrices. *Biomed Imaging Interv J* 5(3):e17
- Lau TK, Bischof WF (1991) Automated detection of breast tumors using the asymmetry approach. *Comput Biomed Res* 24(3):273–295

- Lo SCB, Li H, Wang Y, Kinnard L, Freedman MT (2002) A multiple circular path convolution neural network system for detection of mammographic masses. *IEEE Trans Med Imaging* 21(2):150–158
- Maitra I, Nag S, Bandyopadhyay S (2012) Technique for preprocessing of digital mammogram. *Comput Methods Programs Med* 107(2):175–188
- Moayedi F, Azimifar Z, Boostani R, Ketabi S (2010) Contourlet-based mammography mass classification using the SVM family. *Comput Biol Med* 40(4):373–383
- Narain Ponraj D, Evangelin Jenifer M, Poongodi P, SamuelManoharan J (2011) A survey on the preprocessing techniques of mammogram for the detection of breast cancer. *J Emerg Trends Comput Inform Sci* 2(12):656–664
- Nielsen MA (2015) *Neural networks and deep learning*, vol 25. Determination Press, San Francisco, CA, USA
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9(1):62–66
- Pearson K (1901) On lines and planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci* 2(11):559–572
- Pizer SM, Amburn EP, Austin JD et al (1987) Adaptive histogram equalization and its variations. *Comput Vis Graphics Image Process* 39(3):355–368
- Qiu Y, Yan Sh, Tan M, Cheng S, Liu H, Zheng B (2016) Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study. In: *Proceedings of SPIE 9785, Medical Imaging: Computer-Aided Diagnosis*.
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Networks* 61:85–117
- Setiawan AS, Elysia Wesley J, Purnama Y (2015) Mammogram classification using law's texture energy measure and neural networks. *Procedia Comput Sci* 59:92–97
- Siddiqui MK, Anand M, Siddiqui PK, Anand MKJ, Mehrotra M, Sarangi R, Mathur N (2005) Biomonitoring of organochlorines in women with benign and malignant breast disease. *Environ Res* 98(2):250–257
- Suckling J, et al. (1994) The mammographic image analysis society digital mammogram database. In: *Proceedings of International Workshop on Digital Mammography*. Excerta Medica. International Congress Series 1069, pp 375–378.
- Tavakoli N, Karimi M, Nejati M, Karimi N, Soroushmehr SMR, Samavi S, Najarian K (2017) Abnormality detection of mammograms by discriminative dictionary learning on DSIFT descriptors. 39th International Conference of the IEEE Engineering in Medicine and Biology Conference (EMBC), JeJu Island, S. Korea, July, pp 1740–1743
- The Mini-MIAS Database of Mammograms. <http://peipa.essex.ac.uk/pix/mias> Accessed 2019.
- Tosin A, Morufat A, Omotayo O, Bolanle W, Olusayo O, Olatunde O (2018) Curvelet transform-local binary pattern feature extraction technique for mass detection and classification in digital mammogram. *CJAST* 28(3):1–15
- Wajid SK, Hussain A (2015) Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. *Expert Syst Appl* 42(20):6990–6999
- Zeiler MD (2012) ADADELTA: an adaptive learning rate method. [arXiv:1212.5701](https://arxiv.org/abs/1212.5701)
- Zhang YD, Pan C, Chen X, Wang F (2018) Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. *J Comput Sci* 27:57–68

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