A Deep Learning Approach for Breast Cancer Mass Detection

Wael E.Fathy[1], Amr S. Ghoneim[2]

Teaching Assistant[1], Assistant Professor[2]

Department of Computer Science, Faculty of Computers and Information

Helwan University, Cairo, Egypt

Abstract—**Breast cancer is the most widespread type of cancer among women. The diagnosis of breast cancer in its early stages is still a significant problem worldwide. The accurate classification and localization of breast mass help in the early detection of the disease, so in the last few years, a variety of CAD systems are developed to enhance breast cancer classification and localization accuracy, but most of them are fully based on handcrafted feature extraction techniques, which affect its efficiency. Currently, deep learning approaches are able to automatically learn a set of high-level features and consequently, they are achieving remarkable results in object classification and detection tasks. In this paper, the pre-trained ResNet-50 architecture and the Class Activation Map (CAM) technique are employed in breast cancer classification and localization respectively. CAM technique exploits the Convolutional Neural Network (CNN) classifiers with Global Average Pooling (GAP) layer for object localization without any supervised information about its location. According to the experimental results, the proposed approach achieved 96% Area under the Receiver Operating Characteristics (ROC) curve in the classification with 99.8% sensitivity and 82.1% specificity. Furthermore, it is able to localize 93.67% of the masses at an average of 0.122 false positives per image on the Digital Database for Screening Mammography (DDSM) data-set. It is worth noting that the pre-trained CNN is able automatically to learn the most discriminative features in the mammogram, and then fulfills superior results in breast cancer classification (normal or mass). Additionally, CAM exhibits the concrete relation between the mass located in the mammogram and the discriminative features learned by the CNN.**

Keywords—Convolutional Neural Networks (CNNs); breast cancer; Global Average Pooling (GAP); mass classification and localization; Class Activation Map (CAM); Receiver Operating Characteristics Curve (ROC); Deep Learning; Computer Aided Detection And Diagnosis (CAD)

I. INTRODUCTION

Nowadays, breast cancer is the most common and leading cause of death among women. In comparison to other cancer types, breast cancer is considered the second highest level of expected deaths in women with 14% in 2016. Recently, it has represented a serious health problem worldwide with the highest rate of 29% among other kinds of cancer. Moreover, the number of women diagnosed with breast cancer in 2016 reached 246,660 [1].

About 37.3% of the breast cancer cases which are diagnosed could be entirely healed, particularly, in the case of early detection [2]. In Egypt and other Arab countries, there exist 42 cases diagnosed with breast cancer per 100 thousand of the community. Moreover, breast cancer affects women at the age of thirties in these countries [2]. Breast cancer early detection plays a pivotal role in the diagnosis and the treatment options, and it leads to a 5-year survival rate of 97.5%. In contrast, when the diagnosis delayed and cancer spread to other organs, the patient has a 5-year survival rate of only 20.4% [3].

Mammography is currently the most reliable radiological technique for the early detection of breast cancer. Mammographic screening has been proved its effectiveness in reducing breast cancer death rates by 30-70% [4]. It is difficult to interpret the mammogram since lesions detection in it depends on radiologists' level of experience and also on image quality. Breast cancer diagnostic errors are caused by misinterpretations or overlooking of breast cancer signs. Approximately, 52% of the errors caused by misinterpretations while overlooking signs accounted for 43% of missed abnormalities [4]. The increase of abnormalities' detection failures in the mammogram is due to the poor image quality, eye fatigue, or oversight by radiologists [4].

To overcome the problems associated with mammographic screening, double reading and Computer Aided Detection and Diagnosis (CAD) [5] were introduced in order to increase the accuracy of breast cancer detection in its early stages, thus subsequently decreases the number of unnecessary breast biopsies. In the double reading solution [5], two radiologists review the same mammogram and take the decision. Although double reading can lead significantly to increase the sensitivity and effectiveness of screening, the associated high workload and cost make it impractical. Alternatively, CAD solution was introduced. It combines diagnostic imaging with computer science, image processing, pattern recognition, and artificial intelligence technologies [4]. Therefore, CAD is the second pair of eyes for radiologists, so that only one radiologist is needed to read the mammogram rather than two. It reduces the radiologists' work-load and minimizes the cost while improving the sensitivity of breast cancer early detection [6].

On the report of research by Tang et al. [6], CAD increased breast cancer detection by 7.62%. Additionally, Brem et al., [7] indicated that the use of a CAD significantly increasing the radiologist's sensitivity by 21.2% which led to improving breast cancer detection.

In the recent years, a variety of techniques developed to enhance the accuracy of existing CAD systems, but most of

them are thoroughly dependent on pre-processing, segmentation, and handcrafted feature extraction techniques, which affect the efficiency of the CAD systems. Presently, deep learning approaches deliver a great success in solving computer vision and machine learning tasks [8]; they are capable automatically of learning a set of high-level features which consequently promotes the accuracy of the CAD system instead of handcrafted features [9],[10].

Primarily, deep learning was employed to develop and improve the CAD systems for breast cancer detection [11]. So the main objective of this paper is to introduce a deep learning approach to classify and localize breast cancer mass basing on two related stages: the first aims to use the pre-trained ResNet-50 to extract the high-level features representations from the mammogram and classify them into normal or mass. Results then conveyed to the next stage to localize the breast cancer mass using the Class Activation Map (CAM) technique.

II. LITERATURE REVIEW

Numerous CAD systems proposed for detecting and classifying masses in the digital mammograms. The techniques used for developing these CAD systems categorized into two: the first is composed of multiple steps such as pre-processing, segmentation, feature extraction, and classification steps, which entirely based on image processing and traditional machine learning techniques. In contrast, the second category does not employ any feature extraction techniques for detecting the region of interest, but instead, it exploits all information available in the mammogram using the Convolutional Neural Network (CNN) to learn the features.

Campanini et al. [12] proposed a novel featureless approach for mass detection in digital mammograms. It does not apply any feature extraction techniques for the detection of Region of Interest (ROI); however, it exploits all information available in the image. Two Support Vector Machine (SVM) classifiers were used to reduce the false positive rate. A multi-resolution over–complete wavelet representation is applied to codify the image with redundancy information. The vectors of an immense space obtained and provided to the first SVM to identify it as suspicious or not. The second SVM was used to reduce the false positive rate made by the first, and then classify the input into a mass or non-mass regions. Eventually, the suspect regions detected by using a voting strategy. The proposed approach achieved 80% sensitivity with a false positive rate of 1.1 per image on mammograms from the USF-DDSM database.

Si and Jing [13] presented a CAD system to detect and classify breast cancer mass basing on a Twin SVM classifier. Initially, a mammogram image is intensified using a Dyadic Wavelet-based algorithm. After removing the unwanted noise from a given mammogram, ROI is extracted using a segmentation method combining the Dyadic Wavelet information with mathematical morphology. The suspicious regions were segmented based on an optimal threshold value corresponding to the minimum fuzzy entropy. Afterward, features are extracted from the segmented suspect regions employing Gray Level Differences Statistics (GLDS) and Spatial Gray Level Dependence (SGLD) features. Finally, the Twin SVM classifier is trained and tested to classify masses.

The classifier is trained using 100 masses images and tested using another 100 images from the DDSM dataset. The authors reported that the sensitivity of the proposed system is 89.7% with a 0.31 false positive per image.

Eddaoudi et al. [14] proposed a mass detection system using SVM and texture analysis. ROI classification accomplished in three stages: in the first, a pectoral muscle is segmented using an approach based on contour detection using snakes with automatic initialization. During the second stage, ROI is segmented using maxima thresholding and Haralik features calculated from the co-occurrence matrix. In the third one, a SVM classifier is used to detect whether the extracted features are normal or mass. A classification rate is equal to 77% on average. Authors showed that the results were significantly improved, achieving 95% on average, when the classification applied on the pre-segmented mammograms.

Jen and Yu [15] developed a CAD system for detecting abnormal mammograms by using a two-stage classifier, the Abnormal Detection Classifier (ADC) which applies the Principle Component Analysis (PCA) based technique. To overcome the complexities of the ROI detection in mammograms, primary image processing enhancement techniques were used to remove the unwanted noise, nonbreast regions such as the background, and the spectral muscle. Mammogram's image enhancement leads to detect mammogram's abnormal areas more effectively and precisely. After the pre-processing step, the gray level quantization was used to quantize all ROIs in mammograms and then extract a small number of critical features. All extracted features are classified as normal or abnormal using the ADC. Authors reported that after testing the ADC for 322 images, the sensitivity was 88% and specificity was 84% on MIAS database.

Ertosun and Rubin [16] developed a deep learning visual search system for mass classification and localization in mammograms which comprises two modules: the first is a deep learning classifier to classify the whole mammogram image into two classes (mass and nonmass). While the second aims to localize mass(es) in mammogram images using a regional probabilistic approach based on a deep learning network. Authors reported that the system achieves 85% sensitivity in the classification and 85% in the localization of the masses at an average of 0.9 false positives per image.

Jadoon et al. [17] proposed a three-class (normal, malignant, and benign) mammogram classification using the CNN. This work presented two algorithms: the first based on Discrete Wavelet Transform (CNN-DW); the second bases on Curvelet Transform (CNN-CT). The proposed work shows that extracting the features from the mammogram and using them as an input to CNN is more helpful for cancer detection. IRMA data-set was used to evaluate the proposed method and CNN-DW and CNN-CT achieved an accuracy rate of 81.83% and 83.74%, respectively.

III. PROPOSED APPROACH

A. Data-Set

A subset of mammograms from the Digital Database for Screening Mammography (DDSM) database is used to train

and evaluate the proposed approach. DDSM consists of 2620 cases categorized as 695 normal volumes, 141 benign without callback volume, 870 benign volumes and 914 malignant volumes [18]. For each case, four mammograms captured with two separate views: mediolateral oblique (MLO) and craniocaudal (CC) [18]. The description of DDSM contains the ground truth information associated with each mammogram image with suspect lesions. In our experiment, we have selected 1592 mammograms with mass (benign or malignant) and 2340 normal mammograms. The selected set of mammograms varies between the two views of MLO and CC. The selected data-set divided into 2517, 629, 786 mammograms for training, validation and testing sets respectively.

B. Data-Set Pre-Processing

Pre-processing aims at enhancing the performance of the next stages by applying a set of transformations. The objective of the pre-processing step is to eliminate irrelevant noise and unwanted parts in the background of mammograms to prepare the mammogram images [19] and make them convenient to be analysed by the state of the art deep learning architectures which will also enhance the accuracy of mass detection CAD system.

Original mammogram images have many kinds of artifacts such as medical labels which may connect to the breast region in mammogram and unwanted wide area of the black background that can affect the accuracy of CAD [19]. A sequence of pre-processing steps is applied to remove unwanted artifacts associated with mammogram images. Fig. 2 describes in details steps of the pre-processing stage. Each input mammogram image associated with a ground truth image which is a binary image that represents the mass lesion location with ones. The ground truth image has the same size as its input mammogram image as shown in Fig. 1.

Firstly, a morphological erosion operation is applied to the input mammogram with disk structure element has radius 100 to split any artifacts that may connect to the breast region. Afterward, the breast region is segmented using the ST mapping technique proposed in [32] which generates a binary mask that has ones in the breast region and zeros otherwise. To fill holes that may be caused by previously applied erosion operation, the morphological dilation operation with disk structure element that has a radius of 300 applied to the binary mask. The dilated mask is used to segment the breast region in the input mammogram by setting all pixels' values which not located in the white region of the mask to zeros while preserving the values of pixels found in the breast region which determined by the white region of the mask as illustrated in Fig. 2.
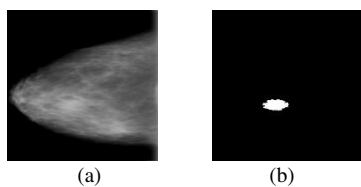

(a)   (b)

Fig. 1. An Example of Input Mammogram Associated with its Ground Truth Image (a) The Input Mammogram (b) The Binary Ground Truth Image Contains a Mass Represented by the White Region.
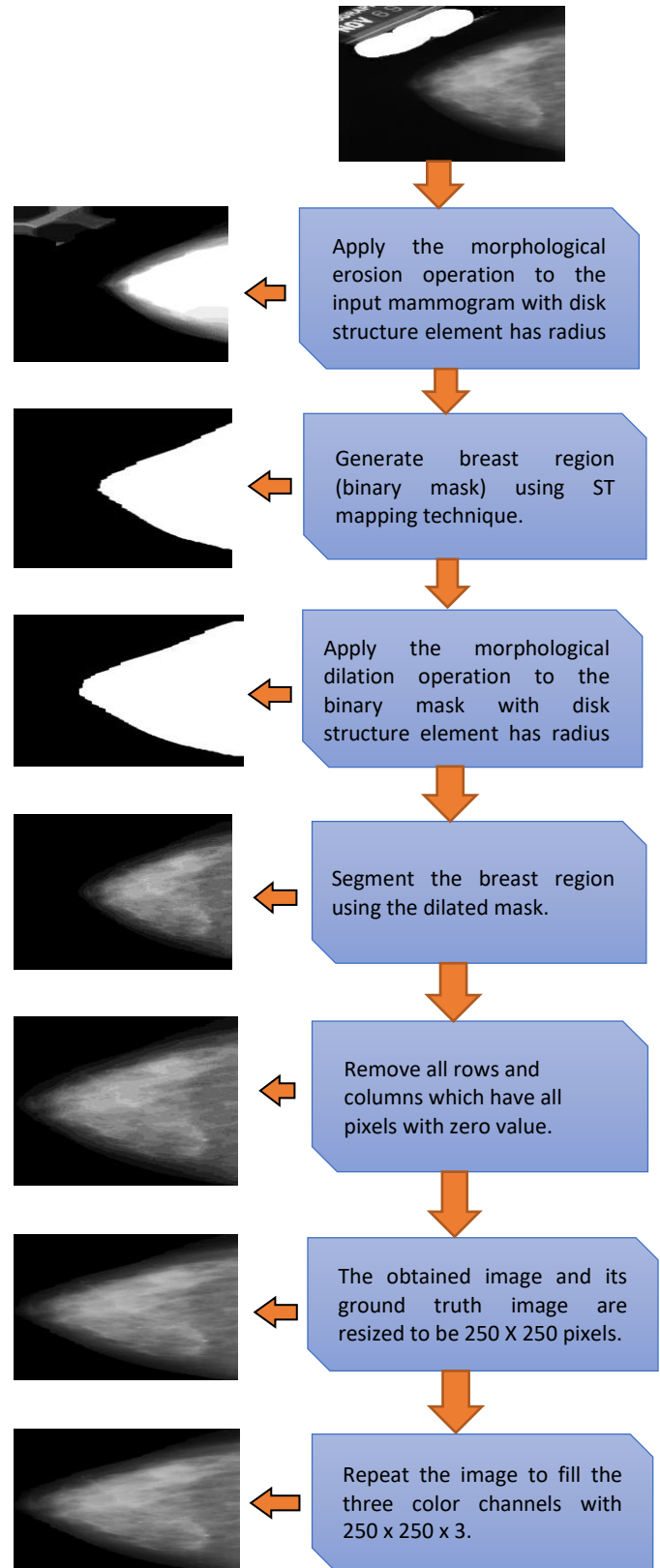

Fig. 2. Pre–Processing Steps.

After applying the previously mentioned steps to the input mammogram, the output is a new image that contains only the breast region represented by the grey area and the background

represented by zeros. The output images usually include columns and rows that have zeros in all of their pixels which do not contain any information about the breast, so that the coordinates of those rows and columns are determined, then they are removed from the image, and its ground truth image respectively as indicated in Fig. 2.

Lastly, the obtained image and its ground truth image are resized to 250 x 250 pixels. Next, the scaled image is repeated to fill the three colour channels with 250 x 250 x 3 pixels to be proper to the deep learning architectures, fit the available memory size and then make the training process as fast as possible. While the output mammogram image has 250 x 250x3 pixels, its ground truth image still has only 250 x 250 binary pixels with ones in mass location and zeros otherwise.

C. Experiment Design

Convolutional Neural Network (CNN / ConvNet) [20] has become the most popular deep learning approach for visual object recognition and classification. In this section, we will designate how to employ the pre-trained ConvNet in the breast cancer mass detection within CAD.

ConvNet [21] composed of a hierarchy of layers inspired by the biological models to transfer information from the lower level to the higher one, introducing more discriminative information in the final representations.

The existence of enough training data enables ConvNet to achieve outstanding results and outperform the traditional hand-crafted methods in object recognition and classification. Large data-sets such as ImageNet and Places contains thousands of images for each class. Provide ConvNet with these datasets to train millions of parameters enables them to achieve extraordinary results [22], [23].

The outstanding ConvNet architectures proposed a few years ago. Meanwhile, there was a noticeable improvement in computational power and optimization methods which facilitated the training of convents and increased its ability to achieve superior results [24]. In our proposed approach, the pre-trained ResNet-50 architecture was selected to compute the CAM for mass localization, as it enables us to compute the CAM without any modification on its original architecture as we will explain in the following subsections.

1) ResNet-50: ResNet [25] has 152-layers network architecture that set new records in classification, detection, and localization problems. It won on ILSVRC 2015 with an incredible error rate of 3.57% top-5 error on ImageNet test set and trained on an 8 GPU machine for two to three weeks. It constructed by the idea of the residual block which makes input *x* go through conv-relu-conv series, for example, assume that *F(x)* is the output of conv-relu-conv series and *x* is the input then:

$$H(x) = f(x) + x \qquad (1)$$

In traditional CNNs, H(x) would be equal to *F(x)*; in this case, *H(x)* called the identity mapping since it computes the transformation of input *x* while concurrently keeping its information [25].

Instead of using fully connected layers in ResNet, the global average pooling layer is proposed to generate only one feature map for each corresponding class and then compute the average for each generated feature map to form a vector fed into *the* softmax layer. The global average pooling layer has many advantages compared to the fully connected layers such as, it has not any parameters to be optimized, so that, overfitting is avoided in that layer. In addition to that, it is robust to the spatial transformation of the input, because it sums spatial changes [25]. Authors construct 5 different ResNet architectures with 18,34,50,101,152 layers respectively [25]. The pre-trained ResNet-50 is selected to extract features from input mammograms and classify them into normal or mass. The activation maps of the last convolutional layer are used to generate the CAM and then localize the most discriminative regions [26]. In the case of the mass class, thus regions usually represent the location of the mass in the mammogram as we will indicate in the results and discussion section.

2) ResNet-50 Training configurations: ResNet-50 architecture trained using Adam optimizer with batch size 16 and learning rate 0.001. The training process finished after 11 epochs using early stopping of patience value of 5. During the training, the best weights saved by the checkpoints on the validation set. Moreover, the pre-trained weights of ResNet-50 fine-tuned and the backpropagation is continued over all layers.

3) Class activation map (CAM) [26]: It is a technique that aims to use image classifier in localization tasks. The idea of CAM is dependent on identifying the most discriminative regions of a specific class without the need for any information about its location during the training. To use the CAM technique for localization, the global average pooling layer added following the last convolutional layer. Global average pooling layer retains the localization details about the object until the closing layer during the classification process. CAM is generated by weighting sum of activation maps in the latest convolutional layer before the global average pooling (i.e., projecting the weights of the classifier on to the activation maps of the last convolutional layer) as in the following equation:

$$CAM_c(x,y) = \sum_{k=1}^{n} W_{k,c} F_k(x,y) \qquad (2)$$

Where c represents the label for a specific class, $F_k(x,y)$ is a k feature map of the last convolutional layer at location (x,y), $W_{k,c}$ is the corresponding weight from k feature map to the class c. CAM_c is the class activation map for the c category. When the generated CAM upsampled to the same size as the input mammogram, the discriminative regions which related to a specific class identified [26].

4) Data augmentation: To avoid the overfitting problem during the training, and then improve the classification accuracy, the following data augmentation methods applied to the training set: random rotation between 0 to 180, horizontal flip, arbitrary height shift (within 0.1 fraction), arbitrary width shift (within 0.1 fraction), vertical flip and arbitrary zoom

(within 0.2 fraction). Thus random transformations [27] artificially increase the training examples, help in avoiding the overfitting and make the model generalize better.

5) Experiment description: In our experiment, we employed the pre-trained restnet-50 architecture to address the problem of the breast cancer mass detection within CAD. Our approach is composed of two phases as indicated in Fig. 3.
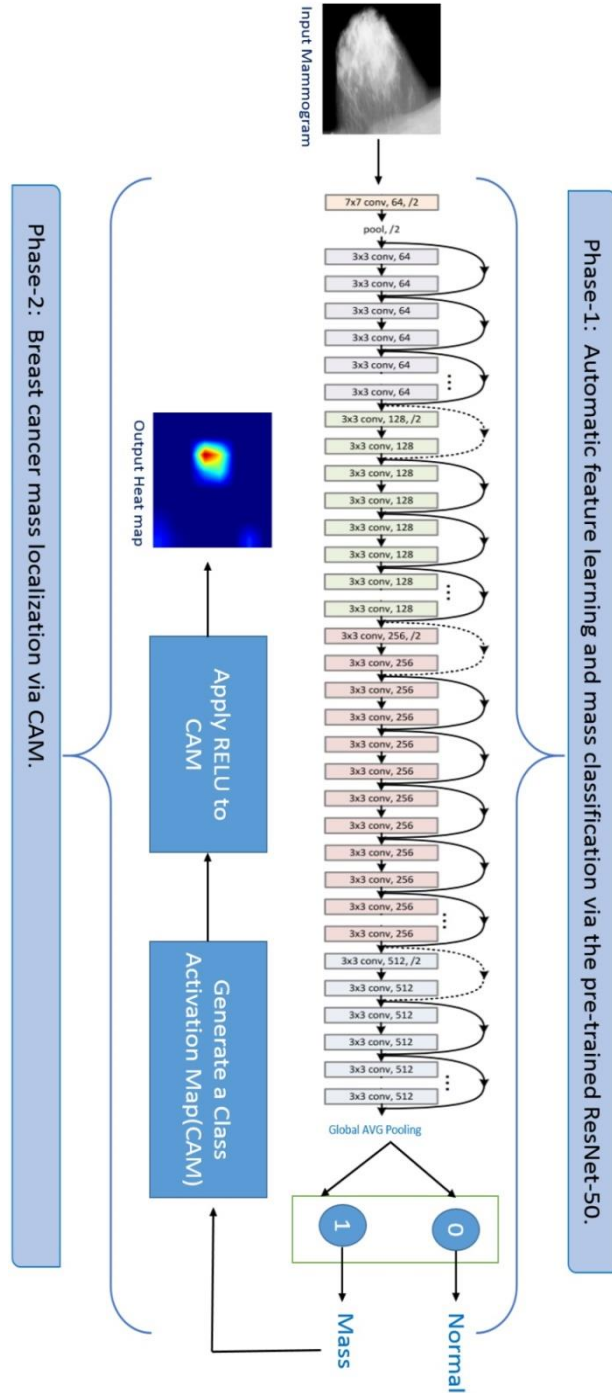
The first phase focused on utilizing the pre-trained RestNet-50 to extract high-level features representations from the mammogram and then classify them into normal or mass class. Furthermore, the second stage focused on substantial breast cancer mass localization via CAM.

To fine-tune the pre-trained model and make it convenient to address the mass detection problem, we added a new layer on the top of the pre-trained model after the global average pooling. This layer acts as a classifier which classifies the input mammogram into two classes (normal or mass) by learning the most informative features about the predicted class. As well, the global average pooling layer preserves the localization details and helps in identifying the most discriminative image regions during the object classification task [26].

According to our approach, if the image classified as a mass class, the CAM will be generated from the last CONV layer. Later, the RELU activation function applied to the generated CAM to threshold it at zero value and then preserving only the positive numbers which hold the crucial mass location details as we will show in the results and discussion section. Lastly, the heat map generated to highlight the most discriminative mass region generated by CAM. Fig. 3 shows the architecture of the proposed approach and describes in details the steps from the input mammogram to the output heat map.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The performance evaluation of the newly developed medical imaging CAD is a significant task which tells us whether the developed system is an improvement over existing systems or not. To evaluate our experiment, ROC and FROC will be used, because they are powerful methods to evaluate medical imaging techniques and compare different proposed approaches [28].

A. Mass Classification

The binary classifier performance is evaluated by the ROC curve. When the classifier classifies a mammogram containing mass as a mass class, this is called a True Positive (TP). Correspondingly, if it classifies a normal mammogram to the class normal, this called True Negative(TN). Terms False Positive (FP) and False Negative (FN) are complements of TN and TP respectively, so that TN + FP = 1 and TP + FN = 1 [29].
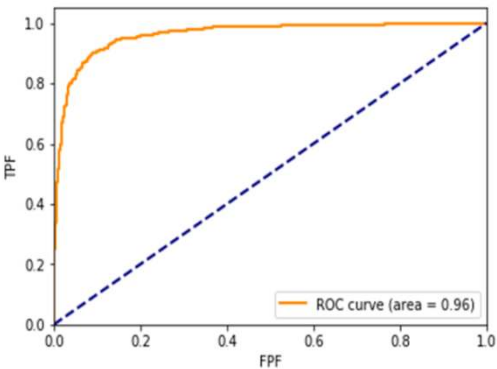

Fig. 3. Architecture of the Proposed Approach.

Fig. 4. ROC Curve of Mass and Normal Classification.

of mass cases which correctly classified as a mass, whereas FPF is the fraction of normal cases which incorrectly classified as a mass [29]. The proposed approach achieves 96% measured by the Area Under ROC Curve (AUC) with 99.8% sensitivity and 82.1% specificity. Fig. 4 shows the ROC for the classification phase.

B. Mass localization

In our approach, the mass localization phase is entirely dependent on the classification of the given mammogram into mass or normal. In case the normal mammogram classified as a normal, it would be considered as true negative, and if classified mass, it would be a false positive. The mammogram deemed a true positive and the mass localized correctly, if and only if the overlapping ratio between its computed CAM and the ground truth mask is 100%. Otherwise, it is considered a false positive. The previous criteria to evaluate mass localization is similar to previous works proposed to localize a mass in the mammogram [30], [31]. In addition to that, the selection of overlapping ration to be 100% aims to measure the ability of CNN architecture with Global Average Pooling Layer to learn the most discriminative features about the object location in the mammogram. Fig. 5 shows the benchmarking results for mass localization via CAM technique.

Ultimately, when the mammogram containing a mass classified as a normal class, it is becoming a false negative. Table 1 shows the confusion matrix that indicates the results for mass localization phase in details.

TABLE I.   THE CONFUSION MATRIX FOR MASS LOCALIZATION RESULTS

Ground Truth/ Predicted	Normal	Mass	Total
Normal	376 (TN)	96 (FP)	472
Mass	20 (FN)	294 (TP)	314
Total	396	390	786

TABLE II.   RESULTS OF OUR APPROACH AND DIFFERENT APPROACHES IN LITERATURE REVIEW

	Classification	Localization
[12]	---	The sensitivity was 80% with a false-positive rate of 1.1 marks per image.
[13]	---	The sensitivity was 89.7% with a 0.31 false positive per image.
[14]	Accuracy was 95%.	---
[15]	the sensitivity was 88% and specificity was 84%.	---
[16]	Accuracy is 85%.	The sensitivity was 85% at an average of 0.9 false positives per image.
[17]	Accuracy is 85%	---
Our approach	96% AUC with 99.8% sensitivity and 82.1% specificity.	The sensitivity was 93.67% at an average of 0.122 false positive rate per image.

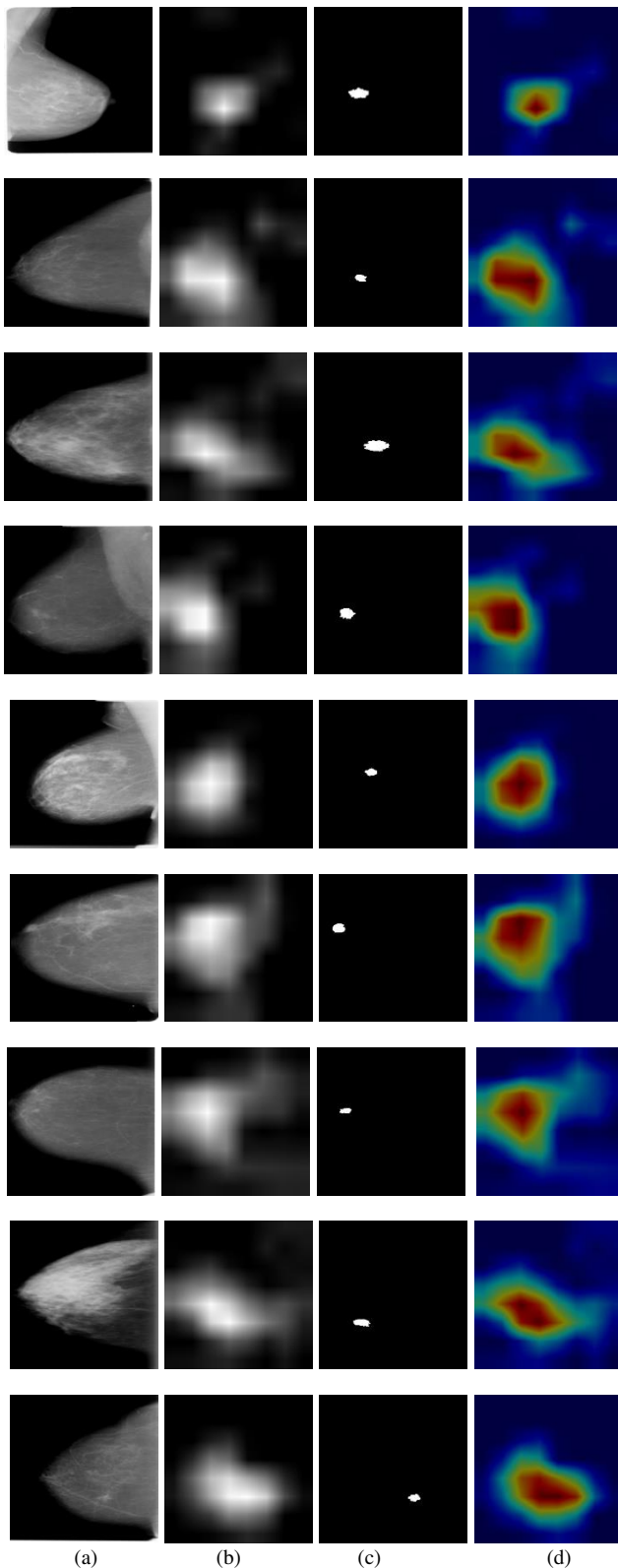(a)   (b)   (c)   (d)

Fig. 5.   Mass Localization Results. (a) Original Mammogram. (b) CAM. (c) Ground Truth Image. (d) Heat Map of the Computed CAM.

ROC curve represents a relation between True Positive Fraction(TPF) or sensitivity on the y-axis and False Positive Fraction(FPF) or 1-specificity on the x-axis. TPF is the fraction

Sensitivity = #TP / #TP + #FN = #true positive / (#true positive + #false negative) =294/314= 0.9367.

Specificity = #TN / #TN + #FP = #true positive / (#true positive + #false negative) =376/472= 0.797.

False Average rate per image = #FP/ Total number mammograms in the test set = 96/786=0.122 per image.

According to the obtained experimental results, our approach is prepared to classify and localize breast cancer masses without using any information about its location. Furthermore, it achieves state of the art result compared to other approaches in the literature review as indicated in Table 2.

Our experiment assures that:

1) The ability of the pre-trained CNN to achieve impressive results in mammogram classification task. Correspondingly, these results can be improved by increasing the training data and train other CNN architectures such as DenseNet.

2) CAM technique is capable of visualizing the class-specific discriminative regions based on the classification results. Furthermore, it provides us with understanding about the concrete relation between the predicted class and its location in the mammogram. Accordingly, the localization results show that 93.7% of masses are fully localized (100%) within the highlighted discriminative regions visualized via CAM. Consequently, CAM can localize mass in the mammogram without presenting any information about its location during the training process as in Fig. 4. Since the mass localization using CAM is wholly dependent on the classification stage, then the mass localization results via CAM can be enhanced by improving the classification results.

V. CONCLUSION

Our work concentrates on classifying and localizing breast cancer mass using the pre-trained ResNet-50 architecture and CAM. The proposed approach composed of two related stages: the first stage aims to classify the mammogram into normal or mass, while the second stage depends on the first to localize mass via CAM.

Experimental results show that the pre-trained ResNet-50 architecture outperforms the traditional techniques in mammogram classification. In addition to that, it shows the ability of CNN to extract the most discriminative features related to a specific class in the mammogram. Additionally, CAM has demonstrated the relation between the discriminative regions of the mammogram and the mass location if the mammogram contains a mass.

In spite of the ability of our approach to localize the mass in the mammogram by computing CAM, the generated CAM is sometimes broader than the mass region in the ground truth image. So we need to apply a specific threshold value to the computed CAM or use a sequence of post-processing steps to reduce it. Accordingly, those notes will be considered in our future work.

REFERENCES

[1] K. D. Miller et al., "Cancer treatment and survivorship statistics, 2016," CA. Cancer J. Clin., vol. 66, no. 4, pp. 271–289, 2016.

[2] G. I. Salama, M. Abdelhalim, and M. A. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers," Breast Cancer (WDBC), vol. 32, no. 569, p. 2, 2012.

[3] A. Jemal et al., "Annual report to the nation on the status of cancer, 1975--2001, with a special feature regarding survival," Cancer, vol. 101, no. 1, pp. 3–27, 2004.

[4] R. M. Rangayyan, F. J. Ayres, and J. E. L. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," J. Franklin Inst., vol. 344, no. 3, pp. 312–348, 2007.

[5] R. M. L. Warren and W. Duffy, "Comparison of single reading with double reading of mammograms, and change in effectiveness with experience," Br. J. Radiol., vol. 68, no. 813, pp. 958–962, 1995.

[6] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," IEEE Trans. Inf. Technol. Biomed., vol. 13, no. 2, pp. 236–251, 2009.

[7] Brem et al., "Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial," Am. J. Roentgenol., vol. 181, no. 3, pp. 687–693.

[8] A. Cruz-Roa et al., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in SPIE medical imaging, 2014, p. 904103.

[9] J. M. Park et al., "Detection and Classification of the Breast Abnormalities in Digital Mammograms via Regional Convolutional Neural Network," pp. 1230–1233, 2017.

[10] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, 2013.

[11] M. A. Jaffar, "Deep Learning based Computer Aided Diagnosis System for Breast Mammograms," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 7, 2017.

[12] R. Campanini et al., "A novel featureless approach to mass detection in digital mammograms based on support vector machines," Phys. Med. Biol., vol. 49, no. 6, p. 961, 2004.

[13] X. Si and L. Jing, "Mass detection in digital mammograms using twin support vector machine-based cad system," in Information Engineering, 2009. ICIE'09. WASE International Conference on, 2009, vol. 1, pp. 240–243.

[14] F. Eddaoudi, F. Regragui, A. Mahmoudi, and N. Lamouri, "Masses detection using SVM classifier based on textures analysis," Appl. Math. Sci., vol. 5, no. 8, pp. 367–379, 2011.

[15] C.-C. Jen and S.-S. Yu, "Automatic detection of abnormal mammograms in mammographic images," Expert Syst. Appl., vol. 42, no. 6, pp. 3048–3055, 2015.

[16] M. G. Ertosun and D. L. Rubin, "Probabilistic visual search for masses within mammography images using deep learning," in 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2015, pp. 1310–1315.

[17] M. M. Jadoon, Q. Zhang, I. U. Haq, S. Butt, and A. Jadoon, "Three-class mammogram classification based on descriptive CNN features," Biomed Res. Int., vol. 2017, 2017.

[18] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in Proceedings of the 5th international workshop on digital mammography, 2000, pp. 212–218.

[19] S. Don and D. Min, "Breast Skin Line Segmentation on Digital Mammogram using Fractal Approach," Indian J. Sci. Technol., vol. 9, no. 31, 2016.

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks.," in CVPR, 2017, vol. 1, no. 2, p. 3.

[21] R. Cichy, A. Khosla, D. Pantazis, A. Torralba, and A. Oliva, "Mapping human visual representations in space and time by neural networks," J. Vis., vol. 15, no. 12, p. 376, 2015.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 248–255.

[23] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in Advances in neural information processing systems, 2014, pp. 487–495.

[24] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in Advances in neural information processing systems, 2007, pp. 153–160.

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

[26] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in Proceedings of the IEEE *Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2921–2929.

[27] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv Prepr. arXiv1712.04621, 2017.

[28] D. P. Chakraborty, "New developments in observer performance methodology in medical imaging," in Seminars in nuclear medicine, 2011, vol. 41, no. 6, pp. 401–418.

[29] X. He and E. Frey, "ROC, LROC, FROC, AFROC: An Alphabet Soup," J. Am. Coll. Radiol., vol. 6, no. 9, pp. 652–655, 2009.

[30] N. D. G. Carneiro and A. P. Bradley, "Automated Mass Detection from Mammograms using Deep Learning and Random Forest."

[31] J. Wei et al., "Computer-aided detection of breast masses on full field digital mammograms," Med. Phys., vol. 32, no. 9, pp. 2827–2838, 2005.

[32] S. Pertuz, C. Julia, and D. Puig, "A novel mammography image representation framework with application to image registration," in Pattern Recognition (ICPR), 2014 22nd International Conference on, 2014, pp. 3292–3297.