



## Breast Cancer Detection, Diagnosis, and Prediction

Asmaa A. Hekal<sup>1</sup>, Ahmed Elnakib<sup>\*2</sup>, Hossam El-Din Moustafa<sup>3</sup>

Electronics and Communications Engineering (ECE) Department, Faculty of Engineering, Mansoura University,  
Mansoura, 35516, Egypt

<sup>1</sup>asmaahekal3@gmail.com

<sup>2</sup>nakib@mans.edu.eg (\*corresponding author)

<sup>3</sup>hossam\_moustafa@ieee.org

### ABSTRACT

The early detection, diagnosis, prediction, and treatment of breast cancer are challenging healthcare problems. This study focuses on outlining the traditional and trending techniques used for breast cancer detection, diagnosis, and prediction, including trending noninvasive, nonionizing, and biomarker genetic techniques. In addition, a Computer Aided Detection (CAD) is introduced to classify benign and malignant tumors in mammograms. This CAD system involves three steps. First, the Region of Interest (ROI) that includes the tumor is identified using a threshold-based method. Second, a deep learning Convolutional Neural Network (CNN) processes the ROI to extract relevant mammogram features. Finally, a Support Vector Machine (SVM) classifier is used to decode two classes of mammogram structures (i.e., Benign (B), and Malignant (M) nodules). The training processes and implementations were carried out using 2800 mammogram images taken from the Curated Breast Imaging Subset of DDSM (CBIS-DDSM). Results have shown that the accuracy of CNN-SVM system achieves 85.1% using AlexNet CNN. Comparison with related work shows the promise of the proposed CAD system.

**Key words:** Breast Cancer, Computer Aided Detection (CAD), Convolutional Neural Networks (CNN), Support Vector Machine (SVM).

### 1. INTRODUCTION

According to the latest report of the international agency for research on cancer, worldwide, breast cancer is the most common cancer among women [1] [2]. In Egypt, cancer is a challenging healthcare problem, especially breast cancer. Breast cancer rates are higher among women in developed countries as well as nearly every region globally [2].

Early detection and classification enable physicians to select the proper treatment option and improves the chance of survival, i.e., early detection can achieve a high cure rate of up to 98% [1]. Standard tests for breast cancer detection depends on age. At early twenties, a breast self-exam every month is recommended. From age 20 to 40, at least one clinical breast exam is recommended each 3 years [1] [2]. Starting from age 40 years, one annual mammogram is highly recommended [1] [2]. These standard tests always represent a challenge in developed countries, due to the relatively high cost and the non-awareness among women societies.

Current mammogram techniques involves the exposure to x-rays, which is an ionized radiation. Ionized radiation causes the atoms to turn to be ions, freeing one or more electrons from the atom. The highly exposure to ionized radiation is unsafe, and may causes mutation and/or cancer. Therefore, exploring new portable, safe, cheap, self-examine technologies is the focus of many recent research papers. Alternatively, the biopsy, which is the gold standard for breast cancer diagnosis, is an invasive technique, which is painful and may cause bleeding. This paper focuses on summarizing the breast cancer traditional and trending techniques for detection, diagnosis, and prediction, including trending noninvasive, nonionizing, and biomarker genetic techniques. In addition, a preliminary system for breast cancer detection is proposed.

The rest of this paper is organized as follows. Section 2 describes the traditional and trending techniques for detection, diagnosis, and prediction. Section 3 presents the proposed system for classifying benign and malignant tumors. Section 4 explores the results and the proposed system findings. Finally, conclusion and future work are illustrated in section 5.

### 2. TRADITIONAL AND TRENDING TECHNIQUES FOR BREAST CANCER DETECTION, DIAGNOSIS, AND PREDICTION

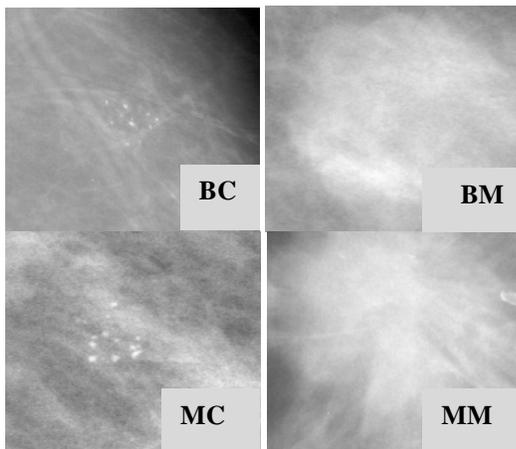
Different techniques are presented for breast cancer detection and diagnosis. The traditional and standard clinical techniques for detection and diagnosis are the mammograms as a

noninvasive imaging technique for detection and following up and the biopsy as a gold standard for diagnosis. This section presents the trends on each method. In addition, genetic biomarkers methods are introduced along with other trending noninvasive and nonionized research methods.

**2.1 Trends on mammogram analysis**

A mammogram intensity values change due to the different X-ray attenuation on the different breast tissues [3]. Figure 1 represents samples of the mammograms of the different types of breast benign and malignant structures, i.e., masses (Benign Mass (BM) and Malignant Mass (MM)) and calcifications (Benign Calcification (BM) and Malignant Calcification (MM)). Different analysis techniques have been presented to detect and classify breast nodules. The traditional methods include the extraction of handcrafted features for classification. On the other hand, deep learning methods, specially the Convolutional Neural Networks (CNNs) represent the state-of-the-art trend for detection, diagnosis, and classification of breast cancer nodules. Deep learning involve the training of two types of layers, convolutional layers and fully connected layers. The deep network involve many convolutional layers in order to extract low-level and high-level features for detection and/or classification. An example of deep CNN network, namely Alex network, is provided in Figure 2.

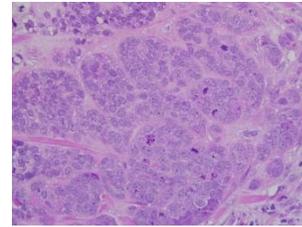
Recently, 3D mammogram is a focus of current research. More specifically, Digital Breast Tomosynthesis (DBT) has been presented in [4], as an alternative to traditional 2D mammogram. However, it is still under research.



**Figure 1:** Types of Mammograms images Benign Calcification (BC), Benign Mass (BM), Malignant Calcification (MC), and Malignant Mass (MM).

**2.2 Histological Image (microscopic images) analysis**

In breast histological imaging, a biopsy sample is prepared by staining it to visualize various tissue components. A standard staining protocol, e.g., using Hematoxylin and Eosin (H&E), is used to selectively stain the blue and pink nuclear structures of the cytoplasm (see Figure 3). Recently, the use of deep learning is the state-of-the-art to analyze the histological images [3][6].



**Figure 3:** Sample of histological Image

**2.3 Biomarkers (genetic analysis)**

Biomarkers are molecules that indicates normal or abnormal conditions or diseases. For cancer, they can be found in all body tissues (the blood, stool, urine, tumor or other tissues).. To characterize cancer signatures, biomarkers are used as tools of molecular biology and are useful for treatment.

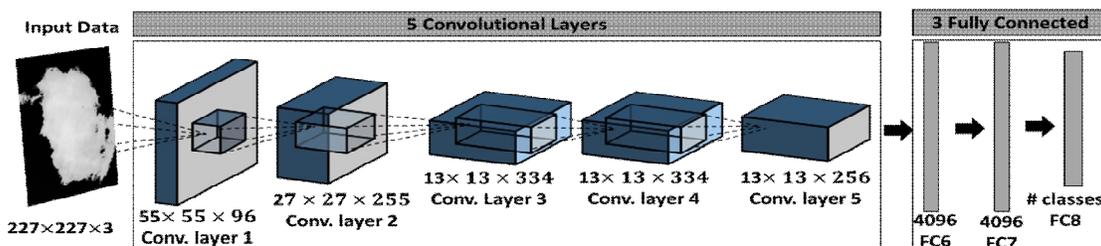
Their applications include the diagnosis, the prognosis, the prediction, and the treatment (see Figure 4).



**Applications of biomarkers**

- Classification of breast cancer
- Determination of the optimal drug and dose
- Prediction of breast cancer and if it can return

**Figure 4:** Applications of biomarkers



**Figure 2:** Deep learning Alex CNN model

**Noninvasive and nonionized techniques**

2.4

The problem with the mammogram is that it uses an ionized radiation, which is harmful and may lead to mutation or cancer, especially for repeated exposure. On the other hand, the biopsy is invasive and is painful and may cause bleeding. Therefore, recent research explores other methodologies for breast cancer detection, such as ultra-wide band (UWB) microwave imaging [7], radar-based microwave imaging [8], and imaging systems at millimeter-waves [9], thermoacoustic (TA) scanner. But these technologies are still under research and development.

### 2.5 Hybrid Techniques

Hybrid techniques involve the integration of two methods such as mammogram/histology, biomarkers/mammogram, etc. The integration of the features extracted from the hybrid methods can significantly improve the accuracy of detection, and or diagnosis [3].

## 3. PROPOSED CAD SYSTEM

In this paper, a deep Computer Aided Diagnostic (CAD) system is presented to classify breast nodules into Benign (B) or Malignant (M). The proposed CAD system is composed of three processing stages (see Figure 5). In the first stage, the ROIs are identified based on a thresholding method. In the second stage, a transfer learning is applied using AlexNet architecture, shown on Figure 2 [10]. Finally, a classification stage is used to classify tumor images. This section illustrate each of these stages.

### 3.1 Collected Database

To test the proposed system, the Curated Breast Image Subset of Digital Database for Screening Mammography (CBIS-DDSM) is used [11]. Examples of the CBIS-DDSM are shown on Fig. 1. In our experiments, the number of images of B nodules is 1400, and of M nodules is 1400, with a total number of the mammogram images of 2800. Image type is converted from DICOM format to PNG format. Figure 6 shows typical samples of CBIS-DDSM (i.e., B, and M), where nodules are outlined using black circles.

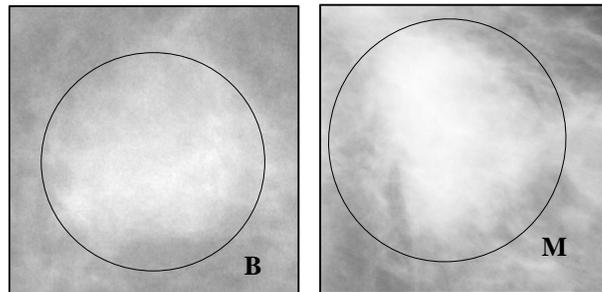


Figure 6: Typical Benign (B) and Malignant (M) samples for the mammograms of the CBIS-DDSM database, where the tumor is outlined using black circles

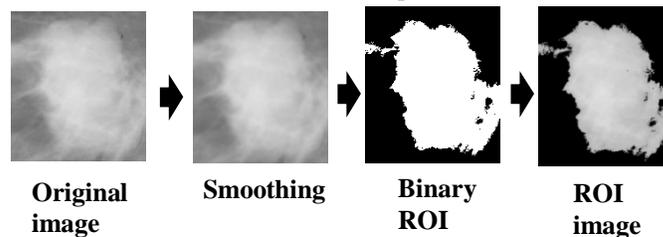
### 3.2 Extracting Tumors

The first stage in the proposed CAD system aims at extracting ROI to be processed by the CNN model. To achieve this goal, a Gaussian smoothing is applied to the database images, and the data base images are normalized between 0 and 1. Then all pixels less than the value of a threshold are classified as background, and other pixels as the required ROI (see Figure 7). By experimentations, the threshold value is selected to be 0.7.

Figure 7: Extracting ROI image by processing the Gaussian filtered image using 0.7 thresholding.

### 3.3 Deep-learning architecture

The ROI images are resized to  $227 \times 227$ , to be suitable for the standard size of the AlexNet CNN input. There are three main



types of layers used to build CNN architectures; CNN architectures consist of three main types of layers; (1) convolutional layer, (2) pooling layer, and (3) Fully Connected (FC) layer [12][13][14].

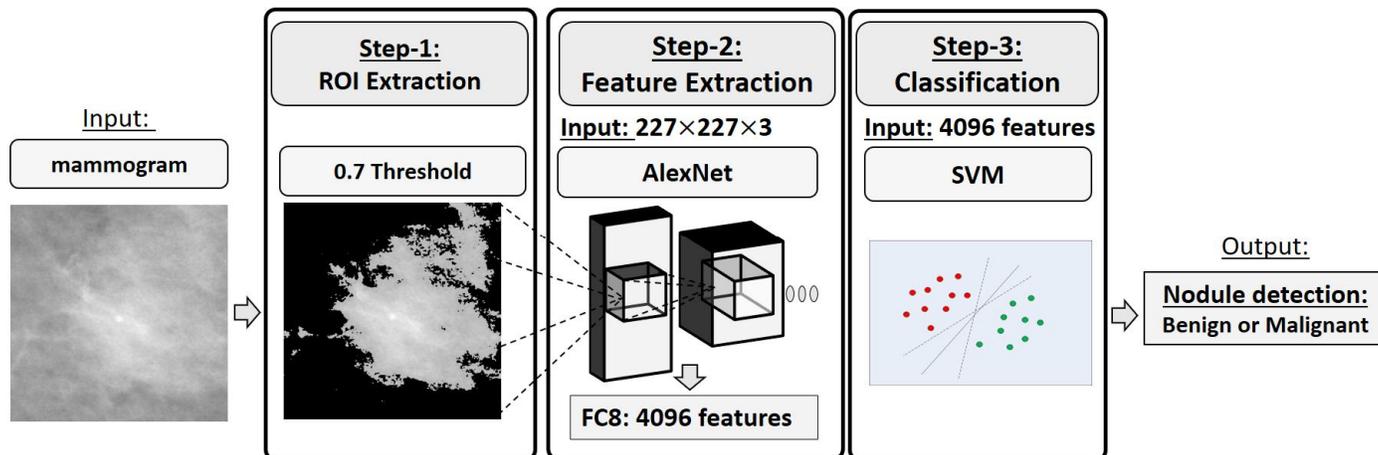


Figure5: The proposed CAD system for breast cancer detection

AlexNet contain eight layers; five convolutional layers, three pooling layers, and three fully connected layers [10] [15]. The AlexNet CNN architecture is shown in Figure. 2. We apply transfer learning [13][14] by keeping the pre-trained convolutional layers and the first two FC layers (FC6 and FC7) as is, and only the last fully connected layer of the pre-trained models (FC8 layer in AlexNet) is replaced by a binary SVM classifier.

### 3.4 Classification

SVM is a supervised machine learning algorithm that sorts data in categories. The idea of SVM is to formulate a computationally efficient way of learning by separating hyper planes in a high dimensional feature space [16]. In this step, ROI images are classified as B or M according to the deep extracted features. The input of the classifier is the vectors of activities FC7 in AlexNet and the output is the class of image.

### 3.5 Performance evaluation

Sensitivity, Specificity and Accuracy are used to quantify the performance of the proposed CAD system [17]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

where *TP* is the true positive, *TN* is the true negative, *FP* is the false positive, and *FN* is the false negative.

## 4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed system sensitivity, specificity and accuracy have been calculated for using AlexNet, and the SVM classifiers breast cancer system. This section illustrates, in details, the experimental setup, results, and related discussions

### 4.1 Experimentation setting

In this work, we used CBIS-DDSM mammogram dataset to verify the proposed methods using MATLAB 2018a. The proposed CAD system is trained and tested on 70% and 30%, respectively, selected randomly from the whole dataset.

### 4.2 Quantitative results

Table 1 compare results using the Accuracy, Sensitivity, and Specificity metrics between three different thresholds to extract the ROI using AlexNet model and the SVM classifier. The best result is achieved when using 0.7 threshold, with an accuracy of 0.851. Without step one of the proposed system (ALexNet+SVM), the accuracy dropped to accuracy 0.65, (see Table 2)

**Table 1:** Accuracy, Sensitivity and Specificity in Proposed system

threshold	Accuracy	Sensitivity	Specificity
0.6	0.830	0.82	0.84
0.7	<b>0.851</b>	<b>0.86</b>	<b>0.84</b>
0.8	0.820	0.81	0.82

Table 2 shows the comparison between the proposed CAD system with different thresholding and the Yi et al. [18] method. The proposed system accuracy is slightly better than Yi et al. [18] method. However, the utilized AlexNet is of much smaller size and less complexity then the Googlenet used by Yi et al. [18].

**Table 2:** Comparison results of Accuracy between Yi [18] and our proposed system

Method	Database	CNN	Accuracy
No threshold	CBIS-DDSM	Alexnet	0.650
0.8 threshold	CBIS-DDSM	Alexnet	0.820
0.6 threshold	CBIS-DDSM	Alexnet	0.840
Yi [18]	DDSM	Googlenet	0.850
Proposed System	CBIS-DDSM	Alexnet	0.851

## 5. CONCLUSION

In this work, the traditional and trending techniques for detection, diagnosis, and prediction are illustrated. In addition, a Computer Aided Detection (CAD) for early detect breast cancer is presented based on deep learning. The CAD system involve three steps: thresholding, deep analysis, and SVM classification. Comparison with related works show the promise of the proposed CAD system. In the future, other CNN models will be investigated. In addition, ensemble deep networks will be considered.

## REFERENCES

- [1] "WHO | Breast cancer," <https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/> May 2020.
- [2] Z. Momenimovahed, and H. Salehiniya. **Epidemiological characteristics of and risk factors for breast cancer in the world**, *Breast Cancer: Targets and Therapy*, Vol. 11, pp. 151–164, Apr. 2019.
- [3] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, and R. Zwigelaar. **Deep learning in mammography and breast histology, an overview and future trends**, *Medical Image Analysis*, Vol. 47, pp. 45–67, July 2018.
- [4] M. Wu, S. Yoon, E. Solomon, J. Star-Lack, N. Pelc, and R. Fahrig. **Digital Tomosynthesis System Geometry Analysis Using Convolution-Based**

- Blur-and-Add (BAA) Model**, *IEEE Trans. on Medical Imaging*, Vol. 35, pp. 131–143, July 2016.
- [5] H. Greenspan, B. Van Ginneken, and R. Summers. **Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique**, *IEEE Trans. on Medical Imaging*, Vol. 35, pp. 1153–1159, 2015
- [6] C. Zhao, and J. Kanicki. **Task-Based Modeling of a 5k Ultra-High-Resolution Medical Imaging System for Digital Breast Tomosynthesis**, *IEEE Trans. on Medical Imaging*, Vol. 36, pp. 1820–1831, Apr. 2017.
- [7] Q. Li, X. Xiao, L. Wang, H. Song, H. Kono, P. Liu, H. Lu, and T. Kikkawa. **Direct Extraction of Tumor Response Based on Ensemble Empirical Mode Decomposition for Image Reconstruction of Early Breast Cancer Detection by UWB**, *IEEE Trans. on Biomedical Circuits and Systems*, Vol. 9, pp. 710–724, Nov. 2015.
- [8] S. Meo, P. Espín-López, A. Martellosio, M. Pasian, G. Matrone, M. Bozzi, G. Magenes, A. Mazzanti, L. Perregrini, F. Svelto, P. Eugene Summers, G. Renne, L. Preda, and M. Bellomi. **On the Feasibility of Breast Cancer Imaging Systems at Millimeter-Waves Frequencies**, *IEEE Trans. on Microwave Theory and Techniques*, Vol. 65, pp. 1795–1806, Mar. 2017.
- [9] D. Byrne, M. Sarafianou, and I. Craddock. **Compound Radar Approach for Breast Imaging**, *IEEE Trans. on Biomedical Engineering*, Vol. 64, pp. 40–51, Mar. 2017.
- [10] A. Krizhevsky, I. Sutskever, and G. Hinton. **Imagenet classification with deep convolutional neural networks**, *Advances in neural information processing systems*, Vol. 60, pp. 1097–1105, May 2017.
- [11] R. Lee, F. Gimenez, A. Hoogi, and D. Rubin. **Curated breast imaging subset of DDSM**, *The Cancer Imaging Archive*, Vol. 8, pp. 2016.
- [12] Y. LeCun, K. Kavukcuoglu, and C. Farabet. **Convolutional networks and applications in vision**, *Proceedings of 2010 IEEE international symposium on circuits and systems*, pp. 253–256, Aug. 2010.
- [13] A. Anagnostis, G. Asiminari, E. Papageorgiou, and D. Bochtis. **A Convolutional Neural Networks Based Method for Anthracnose Infected Walnut Tree Leaves Identification**, *Applied Sciences*, Vol. 10, pp. 469, Jan. 2020.
- [14] T. Truong, and H. Pham. **Breast Cancer Histopathological Image Classification Utilizing Convolutional Neural Network**, *IFMBE Proceeding*, Vol. 69, pp. 531–536, June 2019.
- [15] Z. Wu, C. Shen, and A. Hengel. **Wider or deeper: Revisiting the resnet model for visual recognition**, *Pattern Recognition*, Vol. 90, pp. 119–133, June 2019.
- [16] S. Gunn. **Support vector machines for classification and regression**, *ISIS technical report*, Vol. 14, pp. 5–16, May 1998.
- [17] W. Zhu, N. Zeng, and N. Wang. **Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations**, *Northeast SAS Users Group 2010: Health Care and Life Sciences*, pp. 1–9, 2010.
- [18] D. Yi, R. Sawyer, D. Cohn III, J. Dunnmon, C. Lam, X. Xiao, and D. Rubin. **Optimizing and visualizing deep learning for benign/malignant classification in breast tumors**, *arXiv preprint arXiv:1705.06362*, May 2017.