
M. Al Sharkawy, M. Sharkas, and D. Ragab

Department of Electronics and Communications Engineering
Arab Academy for Science, Technology, and Maritime Transport (AASTMT), Alexandria, Egypt
msharkawy@aast.edu, msharkas@aast.edu, and dina.ragab@gmail.com

Abstract— Breast cancer is one of the most common kinds of cancer, as well as the leading cause of decease among women. Early detection and diagnosis of breast cancer increases the chances for successful treatment and complete recovery for the patient. Mammography is currently the most sensitive method to detect early breast cancer; however, the magnetic resonance imaging (MRI) is the most attractive alternative to mammogram. Manual readings of mammograms may result in misdiagnosis due to human errors caused by visual fatigue. Computer aided detection systems (CAD) serve as a second opinion for radiologists. A new CAD system for the detection of breast cancer in mammograms is proposed. The discrete wavelet transform (DWT), the contourlet transform, and the principal component analysis (PCA) are all used for feature extraction; while the support vector machine (SVM) is used for classification. The system classifies normal and abnormal tissues in addition to benign and malignant tumors. A further investigation was implemented using electromagnetic waves instead of the classical MRI approach. A breast model was generated and near field data of electromagnetic waves were extracted to detect the abnormalities in the breast, especially the masses.

Index Terms — Contourlet transform, discrete wavelet transform, electromagnetic waves, principal component analysis, and support vector machine.

I. INTRODUCTION

For years, cancer has been one of the biggest threats to human life; it is expected to become the leading cause of death over the next few decades. Breast cancer is one of the most common kinds of cancer, as well as the leading cause of decease among women. In Egypt, breast cancer is the most common cancer among women, representing 18.9% of total cancer cases (35.1% in women and 2.2% in men) [1].

Mammography is currently the most sensitive method to detect early breast cancer; however, the magnetic resonance imaging (MRI) is the most attractive alternative to mammogram. Moreover, electromagnetic radiation attracted many researches as a viable alternative to both mammography and MRI.

Micro-calcifications (MCs) and masses are two important early signs of the disease as shown in Fig. 1. There are other less important signs such as architectural distortion [2]. Manual readings of mammograms may result in misdiagnosis due to human errors caused by visual fatigue. Computer aided detection systems (CAD) serve as a second opinion for radiologists. The main goal of the CAD system is to indicate the abnormalities with great accuracy and reliability.

A number of techniques for detection of the abnormalities have been developed in the past decades. Bhangale et al. [3] used the Gabor filters. Strickland et al. [4] introduced the undecimated wavelet transform by combining sub-bands from multiple transforms.
Wang and Karayiannis [5] used the wavelet based sub-band image decomposition. Bruce and Adhami [6] used the DWT modulus maxima method to extract multi-resolutional features that quantify the mass shapes. Yoshida et al. [7] combined a decimated wavelet transform and supervised learning. El-Toukhy et al. [8] introduced the curvelet transform. Gunawan [9] presented a new method for the detection of abnormalities using wavelet transform based on statistical methods; the detection rate was 96%. Juarez et al. [10] applied the functions db2, db4, db8, and db16 of the Daubechies wavelets family, the abnormalities were detected up to 80% accuracy. Balakumaran et al. [11] used the dyadic wavelet transform, and introduced the fuzzy shell clustering algorithm in order to mark the MCs region. Lizcano et al. [12] used the contourlet transform for feature extraction and the support vector machine (SVM) for classification. Addison et al. [13] used the principal component analysis (PCA) as a feature extraction technique to determine the breast cancer. Rejani and Selvi [14] used the SVM for classification; the sensitivity achieved was 88.75%. Sharkas et al. [15] used the proposed CAD system to detect the MCs and to classify between normal, abnormal tissues in addition to benign and malignant MC tumors and the achieved rate was almost 98%.

The aim of this paper is to detect the abnormalities in the breast using CAD systems. A new CAD system is proposed in which the discrete wavelet transforms (DWT), the contourlet transform, and the PCA are used for feature extraction. Moreover, classifying between normal and abnormal tissues, in addition to benign and malignant lesions is presented using the SVM technique. Furthermore, a breast model is generated using electromagnetic waves and near field data are extracted to detect the abnormalities, especially the masses. The proposed CAD system is applied on all the investigated samples.

The paper is organized as follows: Section II describes the CAD system and discusses the used system; Section III discusses the electromagnetic model; Section IV shows the computed results of the proposed technique; Section V discusses the results; and finally, Section VI concludes the presented work.

II. THE CAD SYSTEM

The steps of the proposed CAD system are illustrated in Fig. 2, which is described in details in the following sub-sections.

A. Image enhancement

In this step, the adaptive histogram equalization (AHE) is used, which is an image processing technique used to improve the contrast in images [16]. A generalization of AHE called contrast limited adaptive histogram equalization (CLAHE) was developed, where the histogram is calculated for the contextual region of the pixel [17].

B. Image segmentation

The region of interest (ROI) is extracted from the original mammogram image shown in Fig. 3.
C. Feature extraction

There are many techniques for the feature extraction step; in this paper, the DWT, the contourlet transform, and the PCA are introduced to detect the masses.

1. The discrete wavelet transform

   DWT technique is discussed by many researchers in [2,10,19,20].

2. The contourlet transform

   The contourlet transform or pyramidal directional filter bank (PDFB) is a combination of a Laplacian pyramid and a directional filter bank (DFB) [21]. Band pass images from the Laplacian pyramid [22] are fed into a DFB so that directional information can be captured. The low frequency component is separated from the directional components. After decimation, the decomposition can be iterated with the same DFB in the low pass band to form a pyramid structure. The contourlet transform provides a multi-scale directional decomposition [23].

   Do and Vetterli [24] proposed a PDFB in order to implement the contourlet transform, which is a discrete version of the curvelet transform. The proposed structure is a combination of the Laplacian pyramid and the DFB. Their pyramidal DFB has united two advantages of the two structures, which are multi-resolution and multi-direction. It attempts to separate the low frequency component from the rest directional components and reiterates with the same DFB in the low pass band, forming a pyramid structure. On the other hand, the decomposition solves the problem at low frequency; however, it is still not maximally decimated.

3. Principal component analysis

   Principal component analysis (PCA) is a method that reduces data dimensionality by performing a covariance analysis between factors; it reduces the number of observed variables to a smaller number of principal components, which account for most of the variance of the observed variables. It is used when variables are highly correlated, and it is suitable for data sets in multiple dimensions [25].

   PCA belongs to linear transforms based on the statistical techniques. It provides a powerful tool for data analysis and pattern recognition, which is often used in signal and image processing [25].

   PCA technique can be summarized in the following steps:
   1. Generate the data.
   2. Subtract the mean from each of the data dimensions.
   3. Calculate the covariance matrix.
   4. Calculate the eigenvectors and the eigenvalues of the covariance matrix.
   5. Choose the components and form a feature vector.
   6. Generate the new data set.

D. Classification

In this step, the ROI is classified as either malignant or benign. There are lots of classifiers techniques; among them linear discriminant analysis (LDA), artificial neural network (ANN), binary decision tree, and support vector machine (SVM) [2].

   In this paper, SVM is used; this is due to the fact that it achieved higher classification rates. SVM is a learning tool originated in modern statistical learning theory [26].

   The aim of SVM is to devise a computationally efficient way of learning, separating hyper planes in a high dimensional feature space [26]. There are two cases for SVM; linear SVM and non-linear SVM [27]. The linear
SVM is only used here. This is due to the fact that the investigated data presented in this work is based on a two-dimensional feature space. Thus accepted results were achieved using a non-complex linear SVM technique. For future investigation, a non-linear SVM will be considered.

1. Linear SVM

There are many hyper planes that could classify two sets of data as shown in Fig. 4. The optimum hyper plane that one should choose is the one with the maximum margin. The margin is defined as the width by which the boundary could increase before hitting a data point. The support vectors as shown in Fig. 4 are considered the data points that the margin pushes up (i.e., borderline training sets). Thus, the goal of the SVM is to find the optimum hyper plane that separates clusters of target vectors on the opposing sides of the plane.

![SVM classification including a hyper plane that maximizes the separating margin between two classes](image)

Fig. 4. SVM classification including a hyper plane that maximizes the separating margin between two classes [27].

The SVM formulation is based on the principle of structural risk minimization. It attempts to minimize a bound on the generalization error. Such error is generated from the learning machine on the test data not during the training process. This is instead of minimizing an objective function based on the training samples, such as the mean square error. The SVM is considered robust to work with when dealing with data outside the training set.

### III. THE ELECTROMAGNETIC MODEL

The most common instruments used to detect the breast cancer are mammography and MRI. There are many differences between the two instruments. A MRI uses magnets that emit radio waves to produce a three-dimensional view of the breast and the underlying structures and vessels. Mammography uses low doses of radiation (x-rays) to produce a one-dimensional image of the breast. The breast MRI is a more sensitive test than mammography, especially when given with a contrast medium (i.e., a special dye). The dye makes the cancerous area of the breast appear much brighter, but it may cause many areas of the breast that do not have cancer to appear abnormal, causing an increased number of false positive test results. The breast MRI cannot visualize calcifications. On the other hand, mammography can visualize calcium deposits accurately. The breast MRI may be able to visualize dense breast tissues compared with mammography.

Meanwhile, mammograms and MRI systems do not use ionizing radiation. This leads electromagnetic radiation to become a viable alternative to both mammography and MRI. The microwave (MW) imaging systems offer a promising result in the biomedical applications [28, 29].

The goal of MW imaging is to detect, localize, and characterize the hidden tumors in the breast using electromagnetic waves at microwave frequencies [30, 31]. The MW has low power emission, thus it is harmless. It provides high contrast; this is due to its electrical properties. It also produces a three-dimensional view of the breast. The generated model shown in Fig. 5 is similar to an MRI scan shown in Fig. 6. For implementing the idea and testing the concept an incident plane wave is used to excite the model.

### IV. RESULTS

To verify the proposed method, experiments were performed on the digital database for screening mammography (DDSM) mammogram database [18] using MATLAB. The DDSM database consists of 2620 cases available in 43 volumes. The volumes could be normal, benign, or malignant samples. In this database, the resolution of a mammogram is 50µm/pixel and the gray level depths are 12 bits and 16 bits.
Two hundred and sixty cases were extracted (100 normal cases, 80 benign mass tumors, and 80 malignant mass tumors). First, the samples are enhanced and segmented, and then features are extracted using three ways: DWT shown in Fig. 7, the contourlet transform shown in Fig. 8, and PCA data. The samples go through the SVM technique with linear kernel function for classification.

When classifying normal and mass tissues, the highest classification rate was for the vertical component of the 1st level DWT, which was 93.33%. The approximate component for the 2nd level DWT, achieved 96.67% classification rate as shown in Fig. 9. This proved to be the highest rate when compared to the vertical, horizontal, and diagonal components.

When distinguishing between benign and malignant mass tumors, the contourlet transform features achieved the highest classification rate 92% as shown in Fig. 10.

Tables 1 and 2 illustrate the different rates for the different components of the 1st and 2nd levels DWT in order to distinguish between normal and mass tissues, and benign and malignant tumors, respectively.
Fig. 8. The contourlet transform coefficients for the malignant mass sample shown in Fig. 7 (a).

Fig. 9. SVM classification between normal and mass tissues for the approximate component of the 2nd level DWT features.

Fig. 10. SVM classification between benign and malignant mass tumors for significant contourlet coefficients features.

Table 1: Classification rates for 1st and 2nd levels DWT to classify between normal and mass tissues

<table>
<thead>
<tr>
<th></th>
<th>1st level DWT</th>
<th>2nd level DWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate</td>
<td>83.33%</td>
<td>96.67%</td>
</tr>
<tr>
<td>Vertical</td>
<td>93.33%</td>
<td>83.33%</td>
</tr>
<tr>
<td>Horizontal</td>
<td>56.67%</td>
<td>50%</td>
</tr>
<tr>
<td>Diagonal</td>
<td>56.67%</td>
<td>53.33%</td>
</tr>
</tbody>
</table>

Table 2: Classification rates for 1st and 2nd levels DWT to distinguish between benign and malignant tumors

<table>
<thead>
<tr>
<th></th>
<th>1st level DWT</th>
<th>2nd level DWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate</td>
<td>81.25%</td>
<td>73.75%</td>
</tr>
<tr>
<td>Vertical</td>
<td>86.25%</td>
<td>66.25%</td>
</tr>
<tr>
<td>Horizontal</td>
<td>51.25%</td>
<td>51.25%</td>
</tr>
<tr>
<td>Diagonal</td>
<td>55%</td>
<td>50%</td>
</tr>
</tbody>
</table>

For validation purposes, the receiver operating characteristic (ROC) analysis curve is computed for the vertical component of the 1st level DWT features, which is shown in Fig. 11. The area under the curve equals to 0.917; this means the detection rate is 91.7 %, which is a recognized performance compared to other previous work. This computed value is close to the 86.25 % analytical value for the vertical component of the 1st level DWT. The DWT, contourlet, and PCA feature extraction techniques’ results are discussed in Table 3.

An electromagnetic breast model is generated and simulated using Ansoft high frequency structure stimulated (HFSSTM) software package [33] as shown in Fig. 12. Two-dimensional plane cuts were generated from the model to collect data and test it. A sample of near field data passing through the breast model with a detected tumor is shown in Fig. 13. As previously mentioned in Section III, plane wave excitation is considered for simplicity. It is also used to prove the idea of using electromagnetic waves in conjunction with different image processing techniques for practical detection of tumors as a robust alternative for MRI. Further investigation will be considered for implementing antenna arrays to enhance the tumor detection and provide further hidden information.

Sixty samples (30 benign and 30 malignant) were generated from the model. These samples go through the previous CAD system. The best classification rate was for the approximate and vertical components of the 1st level DWT features.
and the contourlet transform features. Table 4 provides the rates for different components of the 1st and 2nd levels DWT. The results of the feature extraction techniques for the electromagnetic model samples are presented in Table 5.

Table 3: Classification rates for different feature extraction techniques to distinguish between normal and mass tissues, and benign and malignant tumors

<table>
<thead>
<tr>
<th>Feature Extraction Technique</th>
<th>Normal and Mass Tissues</th>
<th>Benign and Malignant Tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWT</td>
<td>96.67%</td>
<td>80.73%</td>
</tr>
<tr>
<td>Contourlet</td>
<td>52%</td>
<td>92%</td>
</tr>
<tr>
<td>PCA</td>
<td>60%</td>
<td>58.75%</td>
</tr>
</tbody>
</table>

Fig. 1. Computed ROC curve for the vertical component of the 1st level DWT for benign and malignant samples.

Table 4: Classification rates for 1st and 2nd levels DWT to classify between benign and malignant mass tumors for the electromagnetic samples

<table>
<thead>
<tr>
<th>Component</th>
<th>1st Level DWT</th>
<th>2nd Level DWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate</td>
<td>99.9%</td>
<td>85.71%</td>
</tr>
<tr>
<td>Vertical</td>
<td>99.9%</td>
<td>92.86%</td>
</tr>
<tr>
<td>Horizontal</td>
<td>50%</td>
<td>64.29%</td>
</tr>
<tr>
<td>Diagonal</td>
<td>64.29%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Fig. 12. Electromagnetic breast model using HFSS™.

V. DISCUSSIONS

For the DDSM samples, the best classification rate achieved when distinguishing between normal and mass tissues was for the DWT features, namely the vertical component of the 1st level.
levelDWT and the approximate component of the 2\textsuperscript{nd} level DWT as shown in Tables 1 and 2. The PCA and the contourlet transform features did not provide promising results, however, the performance of the PCA features proved to be better. Whereas, when distinguishing between benign and malignant mass tumors; the best classification rate was for the contourlet transform features. The DWT features come in the second rank, while the PCA features yielded only 58.75%. This is clear in Table 3.

A benign tumor of radius 4 mm and $\varepsilon_r$ equals to 66 is placed near the fibroglandular tissues. One can easily notice from the electric field distribution shown in Fig. 13 (a) that the tumor is obviously detected due to its high dielectric constant relative to the neighboring tissues. A malignant tumor of $\varepsilon_r$ equals to 20 is detected inside the breast model as shown from the near field data presented in Fig. 13 (b). The tumor in this case is not visible enough; this is due to the fact that its dielectric constant is close to the surrounding tissues.

The classification results of the electromagnetic breast model, shown in Tables 4 and 5, reveal that both the DWT and the contourlet transform have the best classification rate. However, the PCA achieved a weak classification rate. It is clear from the data in Table 5 that the lesions were completely classified (i.e., 99.9 %). This is because the generated samples from the electromagnetic model are not a real extracted data relative to the DDSM database. They were extracted from the proposed HFSS\textsuperscript{TM} model.

VI. CONCLUSIONS

The goal of this work is to detect the masses and classify the tissues in the breast using SVM technique. A new CAD system is introduced, where the DWT, the contourlet transform, and the PCA are used in the feature extraction step. An electromagnetic breast model was generated and tested using the SVM technique and the proposed feature extraction methods and then compared to the DDSM results.

The proposed technique and methods achieved a high classification rates using the DWT and the contourlet transform features. The usage of electromagnetic waves is considered a robust alternative to the MRI for cancer detection. This CAD system can be applied to detect cancer and classify it in any human organ.

ACKNOWLEDGMENT

Great appreciation to Dr. Magda El-Shenawee and her group at The University of Arkansas.

REFERENCES


[32] R. Weisskoff, “MRIImages,” Available online at: int.ch.liv.ac.uk

Mohamed Al Sharkawy was born in Alexandria, Egypt, in 1978. He graduated from the Department of Electrical Engineering, Arab Academy for Science and Technology (AAST), Alexandria, Egypt, on June 2000, and received both, the M.Sc. and the Ph.D. degrees in Electrical Engineering from the University of Mississippi, USA, on October 2003, and December 2006, respectively. From 2002 to 2006, he worked as a research assistant then as a visiting scholar at the Department of Electrical Engineering at the University of Mississippi till June 2007. Currently, he is an Associate Professor at the Electronics and Communications Department at the AAST, where he is currently teaching graduate and undergraduate courses. His research interests include electromagnetic scattering from parallel chiral and metamaterial cylinders and the application of finite difference time and frequency domain techniques for the analysis and design of compact and ultra wideband antennas and microwave devices for wireless communications, radar applications, and breast cancer detection. He has more than 45 papers in international journals and proceedings of international and national symposia and one book by Morgan & Claypool.

Maha Sharkas obtained her B.Sc., M.Sc., and PhD in 1990, 1996 and 2002, respectively from the Electrical Engineering Department, Alexandria University. She is currently a Professor in the Electronics and Communications Department at the AAST. Her research interests are in the field of signal processing, image watermarking, biometrics, biomedical engineering, and pattern recognition.

Dina Ragab graduated from the Arab Academy for Science and Technology (AAST) Faculty of Engineering, Electronics and Communications Department in 2007. She got her Master’s degree in Electrical Engineering in 2011. She is currently a Teacher Assistant at the AAST.