Improved Framework for Breast Cancer Detection using Hybrid Feature Extraction Technique and FFNN

Ibrahim Mohamed Jaber Alamin
Computer Science & Technology
University: Sam Higginbottom Institute of Agriculture Technology and Sciences Allahabad

Dr. W. Jeberson
Associate Professor, Dept. Of Computer Science & Technology, Sam Higginbotom Institute of Agriculture, Technology & Sciences -University, Allahabad

Dr. H K Bajaj
Md (Path), Dcp, Dept.: Health Sciences, Associate Professor, Sam Higginbottom Institute of Agriculture Technology and Sciences –University, Allahabad

***Abstract*—Breast Cancer early detection using terminologies of image processing is suffered from the less accuracy performance in different automated medical tools. To improve the accuracy, still there are many research studies going on different phases such as segmentation, feature extraction, detection, and classification. The proposed framework is consisting of four main steps such as image preprocessing, image segmentation, feature extraction and finally classification. This paper presenting the hybrid and automated image processing based framework for breast cancer detection. For image preprocessing, both Laplacian and average filtering approach is used for smoothing and noise reduction if any. These operations are performed on 256 x 256 sized gray scale image. Output of preprocessing phase is used at efficient segmentation phase. Algorithm is separately designed for preprocessing step with goal of improving the accuracy. Segmentation method contributed for segmentation is nothing but the improved version of region growing technique. Thus breast image segmentation is done by using proposed modified region growing technique. The modified region growing technique overcoming the limitations of orientation as well as intensity. The next step we proposed is feature extraction, for this framework we have proposed to use combination of different types of features such as texture features, gradient features, 2D-DWT features with higher order statistics (HOS). Such hybrid feature set helps to improve the detection accuracy. For last phase, we proposed to use efficient feed forward neural network (FFNN). The comparative study between existing 2D-DWT feature extraction and proposed HOS-2D-DWT based feature extraction methods is proposed.**

Keywords—Breast Cancer; Preprocessing; Segmentation; Region Growing; Noise Removal; Filtering; Orientation; Gradient Magnitude; Higher Order Statistics; FFNN

I. INTRODUCTION

Cancer is the major threat for human being health and its number of patients increasing word wide due to the global warming, even if there are new therapies and treatments proposed by research doctors, but level of cancer defines the ability of its cure. There are different types of cancers from which human being is suffering [male and female]. In this paper we are focusing on breast cancer in women, rest all cancers are out of scope of this paper. Large number of women population is affected by the breast cancer. A different type of reasons causes the breast cancer such as X-Ray [1]. For women's, breast cancer is most common cancer, and it has been increasing since from last decade. The countries like under developed and developed in which breast cancer is commonly observed in females. The estimation of death caused by breast cancer for every year is approximately 40,000 females. This estimation is measured by WHO (world health organization). The world health organization is recognized organization for conducting the research on different cancer diseases. The world health organization also provides the number of breast cancer diagnosis approximately around 200,000. The breast masses evaluation in men is same as in women by considering the mammography [2]. The objective of mammography is conduct the early detection test for breast cancer disease. Mammography is performed based on masses properties as well as micro calcifications. Mammography technology helps to detect the breast cancer before it can happen to individual. But still this approach is not completely accurate. In addition to this, for radiologists it is difficult task to find out the difference between the malignant tumors as well as benign tumors. In mammography, presence of breast cancer is basically reflected. The present approaches considering that recording of image is done over the X-ray film and then that image is interpreted by the medical expertise. However, such approaches are highly vulnerable to visual inspection error and human error. This can be later improved by mammogram images which is of poor quality. The early detection rate is increases based on automated analysis mammogram screening as per the reviews and instigations by different researchers. Another approach is screening mammography which is accurate radiological method currently available for early detection of breast cancer. However as the large number of mammograms needs the analysis, false detections resulted from the radiologists. Therefore, novel techniques for automatic and scalable detection are applicable to overcome such problems. The detection or segmentation of micro-calcification supporting the digital mammogram screening in order divide the clusters as benign or malign [2].

The reason for detection of early breast cancer is that it can helps to cure breast cancer via the proper treatment completely. Such early detection are done by the self-examination process in every month for woman in earlier days. But as discussed in

above paragraph, since from last decade mammography approach is used by many doctors and hospitals for early detection of breast cancer. Micro calcifications and masses characteristics are helps to detect the early cancer for particular individual and hence these are vital factor in detection process. X-ray machines are used to perform the mammography test over the naked upper part of individual. Here in detail mammography is performed as both breasts of women are compressed between the 2 plates with goal of capturing the both photos every breast with help of X-ray pulse. Other well-known methods for early detection breast cancer are CAD (computer aided detection), clinical breast analysis, and blood tests. In order to cure breast cancer completely, it becomes very important to detect it early [3] [4].

CAD becomes interesting area of research since from last decade for early detection of breast cancer to number of researchers as there are number of CAD based automated methods presented by various researchers. The objective of CAD technique is to support radiologists in analysis of breast images by giving the second opinion. The vital goal of CAD is to detect the breast cancer early in women's. Methodology of CAD is based on more than one technique consisting of image preprocessing methods to recognition methods CAD for the detection of CAD based abnormalities in mammograms of breast cancer image. Since from last two decades, number of research groups presented their studies on computer aided diagnosis for early breast cancer detection based on image processing terminologies. CAD takes input as computerized mammographic image which can be generated from the digitally acquired mammogram or traditional film mammogram [4].

The system which is designed based on computer helps to find the abnormal regions of mass, density and calcification in order to diagnose the presence of breast cancer in input image. After detection of this regions, CAD tool highlighting such regions over the original image with aim of further analysis by radiologist. CAD methodology supporting radiologists to make patient management by providing the different recommendations. Since from last 5-7 years, there many advanced CAD systems are proposed by researchers with goal of improving efficiency and accuracy of early breast cancer detection and the objective of assisting the radiologists in interpretation of medical images by providing a second opinion [6] [7]. An important application of CAD is in the diagnosis of breast cancer, which is a common form of cancer diagnosed in women. CAD is an interdisciplinary field, involving elements from basic image processing to advanced machine learning techniques. Therefore use of CAD based detection techniques use is increasing in which image processing concepts are used on input photos from X-ray for automatic detection of breast cancer with its level. CAD system helping to save efforts, time and costs factors for hospitals and doctors. Image processing is nothing but physical method and it is applied in order to convert the breast image signal into the physical image. The image signal is also known as digital image signal, and output of this process is either physical image or its related characteristics. Breast cancer detection is wide range of research in which different researchers preparing their research articles and proposing the new techniques and solutions for breast cancer detection with practical evaluation using the image processing concepts. CAD based techniques are composed of several steps to detect the early detection of breast cancers like acquisition of image, preprocessing of image, segmentation of image, possible feature extraction and then classification for diagnosis. This research paper is contributed by presenting three different phases and algorithms for improving the overall accuracy of breast cancer detection. In this paper contribution is done in four main phases in this work such as preprocessing, image segmentation, feature extraction and classification. Our contributions showing that proposed work improved the detection accuracy as compared to existing approach. In rest of this paper, section II is discussing about the different methods of presented so far on automated breast cancer detection framework. Section III is showing the proposed algorithm, its steps, and inside details for breast image segmentation. Section IV is showing the practical results for this segmentation work on different breast cancer images. Section V presents the conclusion and future work.

II. Related Works

The literature review study of existing methods is considered as one of the important factor that keep the foundation of further system enhancement and development. Therefore, in order to get the information about the existing approaches or systems for breast cancer detection CAD system, a review has been prepared. Below recent ten methods are listed and discussed below for breast cancer detection based on terminologies of CAD system.

- In [4], author Pawar, P.S. et al proposed the novel CAD system architecture for breast cancer detection by implementing back propagation neural network and the authors compared their work with radial basis function network for performance evaluation. The results obtained justifies that back propagation based system performs better for detect breast cancer.
- In [5], author Sameti, M. et al introduced the new image feature extraction technique for the analyzing the screening mammograms retrospectively. This method is taken prior to the detection of a malignant mass in order to detect early breast cancer. For individual mammographic projections of the malignant breast the two specific regions were categorized. The first was for malignant mass subsequently developed and another for similar to region one on the same mammogram. The author employed a stepwise discriminant analysis that exhibited that most of the features could be employed for highly effective classification process of malignant and benign cancer.
- In [6], author Sajjadieh, M.H.S. et al introduced the clutter suppression method referred as DAF/EDF technique which is helps in isolating tumor response from the complete response of tumor and clutter successfully. The presented approach by author is mainly consisting of DFA (data adaptive filter) as well as EDF (envelope detection filter) methods. The benefits of DFA and EDF is that they does not needs any prior training. The implementation approach

followed by authors in which DAF and EDF methods are coupled with TR (time reversal) array method of imaging. This system was analyzed by executing finite difference and time difference (FDTD) electromagnetic simulations depending on MRI (magnetic resonance imaging) of breast data. For microwave detection, this approach was efficient.

- In [7], authors Padmanabhan, S. et al presented the another approach with goal of improving the diagnostic accuracy of early breast cancer using digital mammograms by adopting the simulation tools such as MATLAB with dataset of MIAS. During this work, author introduced the approach for tumor cells detection in order to segment them in various disease phases. Authors also considering the approach of object detection, its recognition and mammograms classification with goal of presenting the difference among abnormal and normal cells. In their study, it was noticed that dense breasts can resulted into more difficult for interpretation with conventional mammograms.
- In [8], author Ben Hamad, N et al presents the study over optimum wavelet approaches and its optimal potential level of decomposition that could provide higher detection accuracy. In the two stage system the author performed multi resolution analysis on the basis of 1D discrete wavelet transform over profiles of micro-calcifications extracted from mammographic images. This was grow up by results validation with the consideration of 2D DWT (discrete wavelet transform) in phases of analysis as well as synthesis over the screening mammograms those are extracted from the MIAS dataset for the micro-calcifications detection.
- In [9], Sridhar, B. et al introduced the approach for automatic image segmentation with goal of tumor detection. The CAD system was designed by authors for the detection of malignant and benign from the images of CT (computed tomography). In their work they employed curvelet transform technique for image segmentation and feature extraction. The main goal of this work was to explore the robustness of curvelet transform which is a multi-scale transform that can represent the edges along curves much more efficiently.
- In [10], author Hussain, M. et al introduced the approach for kernel functions implementation for the early breast cancer diagnosis. The author focused on SVM implementation for the task of classification by considering the various functions of kernel.
- In [11], author Yao-lin Li et al presented function of mixture membership based on the linear distance membership as well as tight density membership. Author focused on efficient classifier design for improving the detection accuracy.

III. METHODOLOGY

The proposed methodology is described in below three contribution points in order to overcome the research challenges of existing automated methods of breast cancer detection in early stage.

- Preprocessing efficiency, in this work we have designed new combined approach for getting better. This preprocessing step is combination of filters like Laplace filter, smooth filter and then binarization and finally smoothing operations. This improves the quality of input raw image more as compared to previous basic preprocessing steps.
- Next contribution of this thesis is used of efficient segmentation method. The existing approach of region growing method is having constraints of orientation and intensity while image segmentation. Therefore the new modified region growing technique is proposed to overcome such constraints by considering both orientation and intensity for efficient image segmentation.
- Feature extraction methods, this is another area in which it is required to have efficient technique in order to get improved recognition accuracy. We proposed the hybrid feature extraction technique which is combination of texture, gradient magnitude, DWT+HOS etc. Another sub contribution is the use of efficient FFNN classifier as compared to existing classifiers.

As per method and flow defined in figure 1 below, three main algorithms are discussed below for preprocessing, segmentation and feature extraction. After that FFNN classifier is applied to get classification accuracy.

Algorithm 1: Preprocessing Algorithm

Step 1: Breast Image Browsing

Step 2: The input raw image needs to be preprocessed.

Step 3: The input image is first resized into 256 * 256 size using the MATLAB function of image resize.

Step 4: 2D conversion, if the input image is 3 dimensional (3D) then it is first converted into 2D, as most of image processing methods are applied on 2D images only. In short, RGB image is converted into gray scale image.

Step 5: Image de-noising is applied by using two filtering techniques mentioned in below steps: Step 5.1: Out_1 = apply Laplacian filter on grayscale image

Step 5.2: Out_2 = apply mean filter on grayscale image

Step 5.3: Out_3 = Out_1 – Out_3

Step 5.4: Out_3 is final preprocessed image

Step 6: Preprocessed breast image

Algorithm 2: Improved Region Growing Segmentation

Input: Out_3 image [Preprocessed Image]

Step 1: Out_3 is preprocessed image from gradient is extracted over X and Y axis in variables OutX and OutY.

Step 2: Combining gradient values using the below equation to get gradient vector Gval.

Gval =[1/(1+(OutX+OutY))]

Step 3: Gval is in radians; hence it is converted to values of degrees in order get orientation information of image pixels.

Step 4: Image Out_3 is divided into grids GRi.

Step 5: Define the threshold values for intensity and orientation in variables Ti and to respectively.

Step 6: for each GRi do

6.1. Compute the histogram Hi of each pixel Pj over grid GRi.

6.2. Searching the frequent histogram of Find the most frequent histogram of GRi grid and referred as FreqH.

6.3. Choose any pixel Pj related to FreqH value, and then assign that pixel information seed point (SP) which is having Ip [Intensity value] and Op [Orientation value].

6.4. Checking the constraint such as intensity and orientation constraints for neighbouring pixel.

6.5. If both a constraint satisfied, then region is grown, else next GRi grid is taken for further processing.

Step 7: Segmented Image

Algorithm 3: Feature extraction algorithm

Input: Segmented Breast Image

Step 1: Extract Texture Features from Input and form feature vector GeF

Step 2: Extract Gradient Features: Gradient and Direction

Step 2.1: Apply mean and standard deviation on gradient

Step 2.2: Apply mean and standard deviation on direction

Step 2.3: Form final 4 features gradient vector called GrF

Step 3: Apply 2D DWT

Step 3.1: Apply mean and standard deviation on LLD

Step 3.2: Apply mean and standard deviation on LHD

Step 3.3: Apply mean and standard deviation on HLD

Step 3.4: Apply mean and standard deviation on HHD

Step 3.5: Form final 8 feature 2D-DWT feature vector called DiF.

Step 4: Apply and Extract Higher order statics using skewness and kurtosis and store all features in vector HoF.

Step 5: Combine GeF, GrF, DiF and HoF to form hybrid feature vectors called CHF.

Output: Feature Vector CHF.

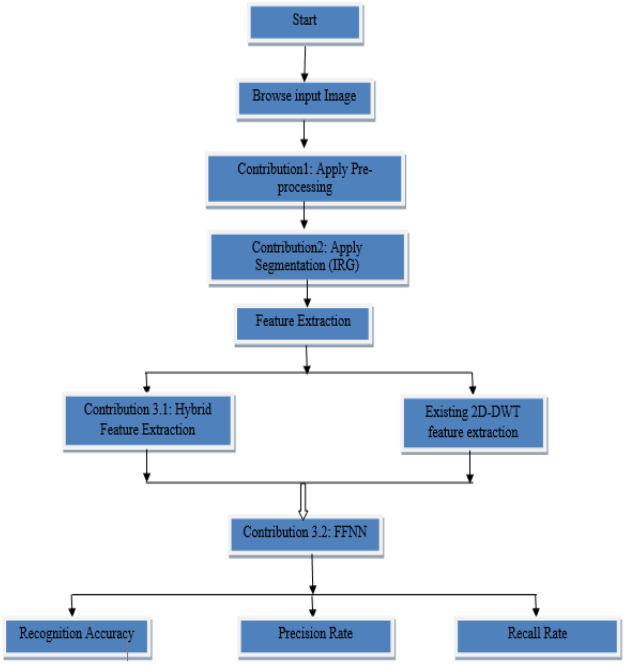

Fig. 1. System Architecture and Flowchart

IV. Experimental Results

A. ***Dataset Information:*** Number of research datasets for breast images is publically available for research studies. For this research two well know datasets such as Mammographic image analysis society (MIAS) and digital database for screening mammography (DDSM) are used. These two datasets are widely used for CAD systems and research works. This dataset we divided into two main classes normal and abnormal with varying number of image samples such as 30, 60, 90, 120, 150 and 180 per class for training and classification purpose.

B. ***Performance Results***

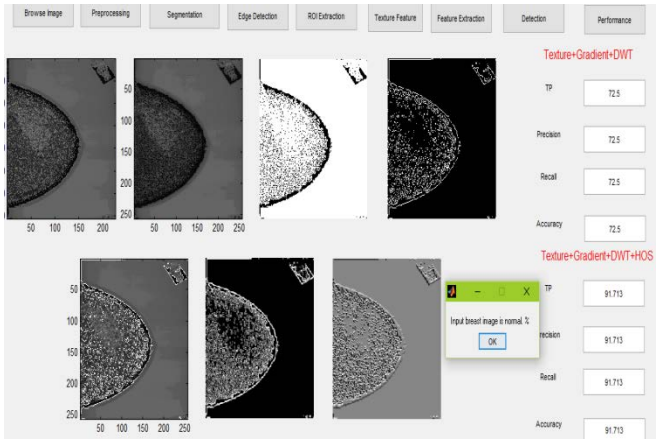

Fig. 2. GUI for proposed framework of automatic detection of breast cancer based on input breast image

Above figure 2 is showing the outputs of all steps and algorithm proposed in this paper such as first task is preprocessing whose output is showing in second window, next is image segmentation using proposed modified region growing technique and its output is showing in window 3 in above figure. After that edge detection, ROI extraction and proposed method of hybrid feature extraction is done. Based on extracted features, classification is done to detect the class of input image whether it is normal or having cancer. Below tables showing the performance analysis for correct classification accuracy and incorrect classification accuracy in between existing and proposed method for varying number of training size. Table 1 showing the comparative study between existing method and proposed method for overall classification accuracy according to varying training sample size.

TABLE I. CLASSIFICATION ACCURACY PERFORMANCE ANALYSIS

Training Size	Existing Method Accuracy (%)	Proposed Method Accuracy (%)
30	72.5	91.71
60	73.1	79.17
90	75.003	81.7
120	70.27	75.71
150	77.57	80.8
180	66.78	81.5

TABLE II. INCORRECT CLASSIFICATION PERFORMANCE ANALYSIS

Training Size	Existing Method Accuracy (%)	Proposed Method Accuracy (%)
30	27.5	8.28
60	26.89	20.82
90	24.99	18.29
120	29.72	24.28
150	22.42	19.19
180	33.21	18.49

From above table 1 and 2, it is clear that proposed method for automatic breast cancer detection is performing better as compared to existing method for all types of training sizes. The performance analysis for correction classification and incorrect classification is depicted in table 1 and 2 respectively. Figure 3 is showing the comparative graph analysis for detection accuracy of 2D-DWT based system and proposed 2D-DWT + HOS feature extraction methods.

V. CONCLUSION AND FUTURE WORK

The goal of this research article is to focus on improving the detection accuracy of CAD technique for breast cancer detection. By considering this objective, this paper presenting the contribution, its framework, flowchart and parameters with simulation environment. Based on this proposed work methods, simulation study is conducted by using publically available research dataset for normal and abnormal breast images of different candidates. The experimental results introduce the goal and main contribution of this research work. The automated computerized breast cancer detection framework presented in this thesis is having maximum accuracy is around 91 % under real time simulation environment which is more as compared to existing or previous methods which are based on SVM-model and 2D-DWT or relevant features extraction techniques. The maximum accuracy of existing methods is around 77 % for correct classification. Therefore, proposed approach of breast cancer detection improving the overall accuracy by 15 % in an average for all sizes of training. The possible future work for this to extend this work by online breast cancer detection system as current system is offline.

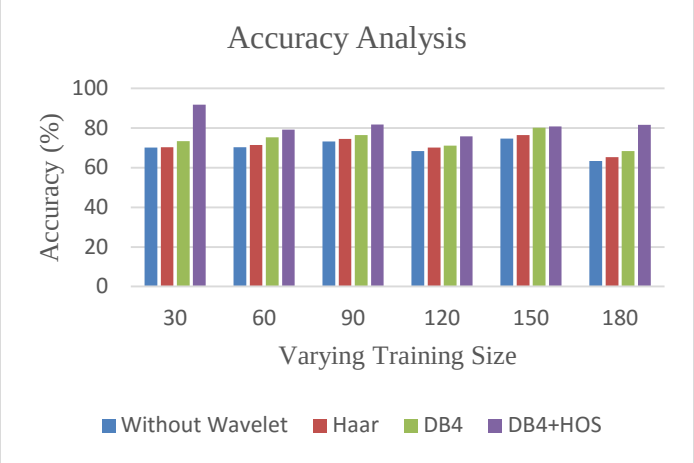

Fig. 3. Comparative analysis of different CAD methods

REFERENCES

[1] Babu, G.A.; Bhukya, S.N.; Kumar, R.S., "Feed forward network with back propagation algorithm for detection of breast cancer," Computer Science & Education (ICCSE), 2013 8th International Conference on 26-28 April 2013, vol., no., pp.181-185.

[2] Oral, C.; Sezgin, H., "Classification of mammograms using multilayer neural network," Electrical, Electronics and Computer Engineering (ELECO), 2010 National Conference on 2-5 Dec. 2010, pp.512-515.

[3] Alias, Norma; Ghani, A.C.A.; Saipan, H.F.; Ramli, N.; Palil, S.Q.M., "Wave equation for early detection of breast cancer growth using MATLAB Distributed Computing," Enabling Science and Nanotechnology (ESciNano), 2012 International Conference on 5-7 Jan. 2012.

[4] Pawar, P.S.; Patil, D.R., "Breast Cancer Detection Using Neural Network Models," Communication Systems and Network Technologies (CSNT), 2013 International Conference on , vol., no., pp.568,572, 6-8 April 2013.

[5] Sameti, M.; Ward, R.K.; Morgan-Parkes, J.; Palcic, B., "Image Feature Extraction in the Last Screening Mammograms Prior to Detection of Breast Cancer," Selected Topics in Signal Processing, IEEE Journal of , vol.3, no.1, pp.46,52, Feb. 2009

[6] Sajjadieh, M.H.S.; Asif, A., "Unsupervised time reversal based microwave imaging for breast cancer detection," Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on 8-11 May 2011, vol., no., pp. 1411- 1415.

[7] Padmanabhan, S.; Sundararajan, R., "Enhanced accuracy of breast cancer detection in digital mammograms using wavelet analysis," Machine Vision and Image Processing (MVIP), 2012 International Conference on 14-15 Dec. 2012 vol., no., pp.153-156.

[8] Ben Hamad, N.; Ellouze, M.; Bouhlel, M.S., "Wavelets investigation for computer aided detection of microcalcification in breast cancer," Multimedia Computing and Systems, 2009. ICMCS '09. International Conference on 2-4 April 2009, vol., no., pp.547-552.

[9] Sridhar, B.; Reddy, K.V.V.S., "Qualitative detection of breast cancer by morphological curvelet transform," Computer Science & Education (ICCSE), 2013 8th International Conference on , vol., no., pp.514,517, 26-28 April 2013.

[10] Hussain, M.; Wajid, S.K.; Elzaart, A.; Berbar, M., "A Comparison of SVM Kernel Functions for Breast Cancer Detection," Computer Graphics, Imaging and Visualization (CGIV), Eighth International Conference on 17-19 Aug. 2011, vol., no., pp.145-150.

[11] Yao-lin Li; Jun Feng,; Yan Ren; Qiu-ping Wang; Bao-ying Chen, "Breast cancer detection based on mixture membership function with MFSVM-FKNN ensemble classifier," Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on 29-31 May 2012, pp.297-301.