

Image Processing & Neural Network Based Breast Cancer Detection

Marwan Abo Zanona¹

¹ Department of Computer Sciences, College of Shari'a and Islamic Studies in Al Ahsaa, Al-Imam Muhammad Ibn Saud Islamic University (IMSIU), Al Ahsaa, Saudi Arabia

Correspondence: Marwan Abo Zanona, College of Shari'a and Islamic Studies in Al Ahsaa, Al-Imam Muhammad Ibn Saud Islamic University (IMSIU), Al Ahsaa, Saudi Arabia. Tel: 966-531-321-145. E-mail: dr.marwanzanona@yahoo.com

Received: February 2, 2019

Accepted: February 29, 2019

Online Published: April 30, 2019

doi:10.5539/cis.v12n2p146

URL: <https://doi.org/10.5539/cis.v12n2p146>

Abstract

A large percentage of cancer patients are breast cancer patients. The main available methodology to examine the breast cancer is the Mammography. It detects the signs of breast cancer as different signs supports the experts' decision. Actually, the Mammography is based on human perception and observations. So, build an AI computerized system will take major role in early signs detection. This paper presents an image processing with aid of artificial neural networks computations for computerized signs detection and exploration of breast cancer. The input material is the mammogram images, and the output helps the pathologists to take a decision. A set of input mammogram images was used for development, testing, and evaluation. The mammographic image will be preprocessed and then the features will be extracted using discrete wavelet transformation with aid of Weiner filtration. A historical data of extracted features were used to train a neural network, while the historical extracted features contains both Cancer and non-Cancer images. The combination of neural network machine learning, and rigid image processing techniques resulted accurate outputs. The methodology and results are showed and discussed later in this paper.

Keywords: neural network, supervised learning, weiner filter, discrete wavelet transform, feature extraction

1. Introduction

The breast cancer has been listed as one of most and main dangerous diseases in the 21st century. Breast cancer represents 23% of the total invasive cancer that female people can hold. That makes most of the world wide organizations and countries put an effort in early discovery of breast cancers. The early discovery improves the chances of early treatment to be successful and to make the patient live normally. In fact, cancer could be detected early by its signs. The signs are being decided by specialist physiotherapist. Breast cancer signs are analyzed, studied, and discovered by using mammography, which is a type of x-rays, MRI, ultrasonic imaging, or even nuclear imaging. But the most common is the x-rays mammography. Mammography is the best cheapest and easier methodology for breast cancer imaging, analysis, signs detection, and physiotherapist's decision making. X-Ray is used to treat tumors before its spread throughout the human body, and help the specialists to identify cracks, infections, injury, and abnormal bones. It also helps in locating cancer inside body areas. It shows very fine features and can access different layers inside human body (Sheshadri & Kandaswamy, 2006; Kother, Arumuga & Sathik, 2011; Ponraj, Jenifer, Poongodi & Manharan, 2011).

Many methodologies and tools are being designed to aid the physiotherapists to make a decision regarding the early signs of breast cancer, as well as other diseases. The secret in mammogram images is its high sensitivity that can reach up to 90%. This makes it the main methodology used to analyze breast diseases. When mammography images show abnormal flows, defects, or cases in clean area, the physiotherapist almost decides to make biopsy procedure to study that abnormal laity, and to decide whether it is cancer or not (Gayathri, Sumathi, & Santhanam, 2013).

Hence, the mammography is the main methodology helping in breast cancer detection, though a double checking should be made by different radiologists and physiotherapists. The accuracy couldn't be achieved without double checking, where it could increase the accuracy by 15% over the single checking. This methodology is time consuming, hard to be achieved, arguing methodology, and it requires many human resources. In addition to that, this methodology depends on human perception, human eye, and knowledge (Gayathri, Sumathi & Santhanam,

2013).

The disadvantages of the conventional methodology will be minimized in this paper, by the adoption of computer aided system and artificial intelligence techniques. Computerizing the conventional methodology will support the decision making using computer software that preprocess the mammographic images, analyze it, and then makes a decision regarding the specific patient imaging results (; Kother, Arumuga & Sathik, 2011).

The presented algorithm contains two steps; the first step is to preprocess the mammographic image using different image processing techniques, including special operations and morphological processes operations. The second is to build machine learning based system that takes different historical mammographic images that describes the status of different cases including real patients and non-patient people, and train the artificially intelligent computational system. The machine learning system that will be adopted here in this paper is the feed forward artificial neural network, with multi-layer nodes design (Rahman, Nesha, Akter & Uddin, 2013).

The raw mammographic images will be filtered to reduce the noise as the first stage of preprocessing algorithm. Edge detection, segmenting ROI (region of interest), and detecting masses are the main steps of image preprocessing phase. The algorithm is based on hybrid techniques to detect the cancer features. The preprocessing of the breast mammographic images aims to crop the breast ROI and make it ready for the next phases of processing algorithm. Feature extraction is being done using discrete wavelet transformation with aid of Weiner filtration to extract the features of interest that loads the cancer features.

The recognition of extracted features (i.e. or actually classification) is based on machine learning using artificial intelligent neural network with feed forward recognition and back propagation learning procedure. In such way, a historical data of both cancer defected and non-cancer defected extracted features of mammographic breast images will be used as historical learning data of the neural network. Then, the network will be able to recognize the cancer features during feed forward cycle.

2. Related Works

Only few researches out of many practical ones resulted in meaningful and useful output. The AI has the main role in the decision making of cancer detection. Hence, the mammography is in fact a type of x-rays that reconstructs an image, the digital mammography is a type of DICOM images that represents the x-rays mammography images. The digital medical images could be analyzed and processed by computer. The AI techniques of breast cancer detection attempts to process and / or analyze the digital mammographic images and support it with decision making engine. The following is a discussion of the main related researches.

Sheshadri and Kandaswamy in (2006) presented a thresholding technique for the ROI (region of interest), then, filtering and preparing it for identification process. A micro classification technique was demonstrated based on mammogram images that were segmented and analyzed using ROI and specific filters. That algorithm was supported by computer based decision making to decide if the micro image is classified as normal or ab-normal behavior (Sheshadri & Kandaswamy, 2006).

Kother in (2011) used wavelet decomposition to analyze the image and enhance it versus noise. That paper presents an enhancement and preprocessing technique (not decision making) for mammogram images in order to enhance physiotherapist decision making based on the raw data (Kother, Arumuga & Sathik, 2011).

In (2012) Ponraj, presented morphological preprocessing of mammogram images to enhance the image clearance, contrast, and noise removal. The main goal was to get higher contrast images. The aim of that paper also was to develop and enhance the raw image in order to support the physiotherapist decision making process (Jenifer, Poongodi & Manharan, 2011).

In (2013) Gayathri, concentrates on decision making rather that preprocessing and enhancement. His research was based on machine learning with different algorithms and approaches to determine if there is a propability of breast cancer or not. Supervised, un-supervised, and semi-supervised learning algorithm was tested. The main result of this research was that the AI machine learning is capable to detect the breast cancer initial signs (Gayathri, Sumathi, & Santhanam, 2013).

Salahat, Sale, and their partners (Salahat, Saleh, Salahat, Sluzeh, Al-qutayri, Mohammad & Elnaggar, 2016) presented a method to detect cancer using MRI set of images. The MRI images are converted to binary and scanned for contrast and equal value regions. The higher contrast area then expected to be an ROI. Those areas are expected to be cancer regions. This patent demonstrated a direct scheme for image processing and finding the extraordinary contrast scheme.

The patent (Yamamichi, Umeki & Akagi) that was presented by Yamamichi, Umeki, and Akagi demonstrated a

method for breast cancer detection by image processing, depending on the skipping areas that is shown in the chest wall of the image by processing the images of both breasts and compare them together to find the extraordinary areas that represent the ROI. This patent was concentrated on building the hardware device rather than the software algorithm.

Fredric C. Mertz, and his partners presented a method for early cancer detection by processing the skin images, (Mertiz, Pina, Chou & Seybold, 2012). The patent was concerned with aligning the features of interest of skin that could be affected by cancer. The features were based on reference image and other patient images. The interrelation between the skin areas is being estimated using the curvature analysis. The difference between register areas of skin is entered to decision making to determine if this is expected to be abnormal skin or not. This patent was concentrated on software algorithm rather than the image hardware.

Patil and Jain (2014) presented an algorithm to detect cancer cells using digital image processing throughout thresholding and watershed. They depended on image preprocessing techniques for image conversion and noise removal, then segmentation of the image to detect ROI. The result was demonstrated in terms of thresholding and watershed transformation technique ([Bhagyashri & Sanjeev, 2014).

In (2017) Kalaivani, Chatterjee, Juyal, and Gupta developed a technique for Lung cancer detection using digital image processing methodology that is supported by artificial intelligent neural network. The authors depended on preprocessing the cancer image with aid of histogram then extracting features using neural network (Kalaivani, Pramit, Shikhar & Rishi, 2017).

This presented paper focuses on both, preprocessing and enhancement of mammography images, and decision making is based on machine learning and artificially intelligent neural networks architecture.

3. Methodology

Figure 1 illustrates the main methodology of image processing algorithm in general. The first stage is to get the digital image acquired by the computer system. The digitizing of that image is to scan it using special optical scan device that converts the analog image to DICOM (digital image and communication on medicine). The DICOM images are high resolution, high contrast, high color depth gray scaled images, which are commonly used in medical applications. The medical application images contain a lot of data and large number of details. The basic detail in medical imaging is the color intensity of the images components / contents. The color intensity is the main data holding all medical reconstructed data of medical devices or machines.

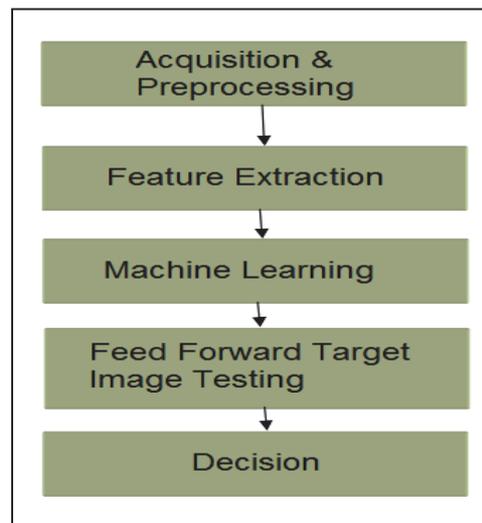


Figure 1. Image processing flow

The large details couldn't be included within 8-bits gray scaled normal digital images, as normal digital imaging techniques. Thus, DICOM images require larger color depth digitizing in order to save the detailed important data of analog medical image. The DICOM images are 16-bits.

Once the digitizing of medical images is done, the raw image will be ready for processing. After preprocessing, the image segmentation process focusing on ROI should be done using rigid algorithm and technique.

The feature extraction follows the segmentation process. It concentrated on extracting the meaningful data (i.e. features) that hold information or data regarding specific problem. The extracted features as well as the feature extraction process is problem dependent and could be achieved by different techniques and methodologies depending on what kind of features the problem concentrates on to solve or to make a decision.

Features will be extracted using discrete wavelet transformation and Weiner filter, where the Weiner will be firstly applied to the cropped breast image that represents the ROI, then the output of this filtration process will undergo the transformation of discrete wavelet transformation (DWT). The extracted features then will load the information of cancer or not. This information will be recognized by machine learning via neural network. Machine learning will be completed using historical data of cancer defected images and non-cancer images that both were entered to the neural network by backpropagation learning algorithm concatenated with the historical decision that is related to that data. Once learning is completed, the neural network will be ready to be used in evolution or decision making in feed forward cycle of the network.

Evolution or decision making process is the final step that concentrates on making a specific decision for specific problem. In this paper, the decision making is based on neural network, the feature extraction process is divided into two phases, image processing and artificial neural networks.

The use of Weiner filter for removing unwanted noise will be explained later in this section.

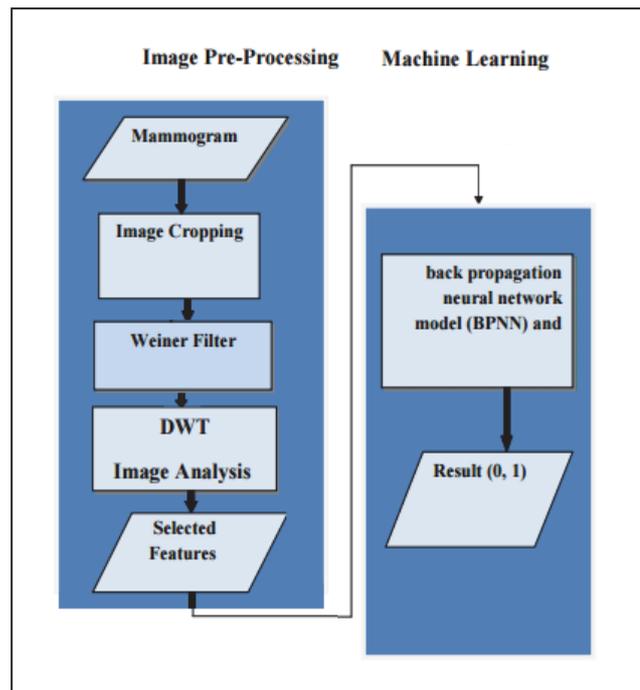


Figure 2. Algorithm's flow chart

Figure 2 illustrates the detailed flow chart of the proposed algorithm, illustrating the two phases of the algorithm, image processing and machine learning.

Cropping the image deletes the un-needed part of the raw image. Hence the input image raw data is the mammographic image. Surely it may contain larger area than needed of the human body, or another object in the image that is not related to the patient. So, cropping the image makes the output image focus only on the reasonable object within the raw image. The margins of the image also is removed after cropping. Figure 3 shows a sample cropped raw mammographic image.

The main noise of the raw image is represented as blurring noise. The noise removal presented in this paper is based on Weiner filter. Weiner filter could be used in different modes, the de-blurring was designed. Figure 4 illustrated the use of Weiner filter, where figure 4.a illustrates blurred image before filtration, and figure 4.b illustrate filtered image, without blurring.

Weiner filter extracts features that are considered time-invariant features (which is important for cancer detection), that is the main reason explaining why to use Weiner filter. Weiner filter performed good results in preprocessed

frequency domain signal.

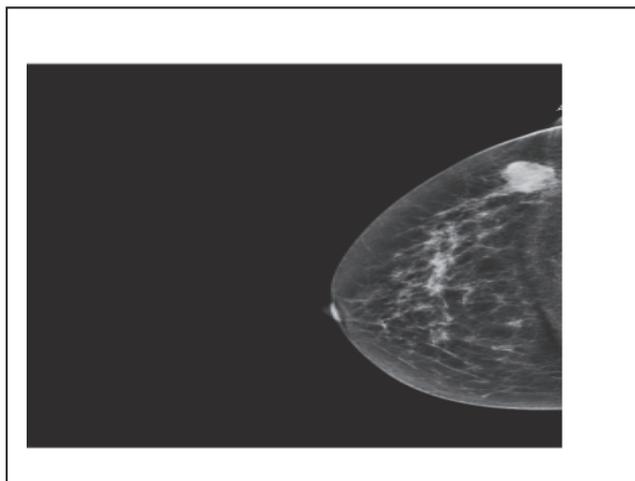


Figure 3. Sample cropped raw mammographic image

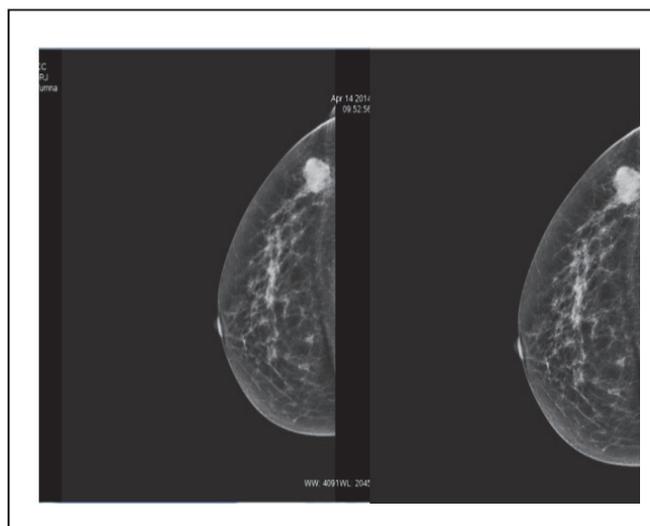


Figure 4a. original blurred image Figure 4b. filtered image (un-blurred)

The image then is converted into black and white to minimize features. The conversion from gray to black and white, or banalization is based on thresholding. In this paper, banalization is done based on the histogram analysis, where the maximum value of the histogram is selected as the threshold value of the gray to black and white conversion.

Figure 5 shows the banalization result of the sample image under processing. Figure 5.a demonstrates the gray image and figure 5.b the binary image. The resulted binary image contains more specific information regarding the internal structure and additions of the breast. This removes many unwanted details of the image and keeps the possible cancer points to be processed further ahead.

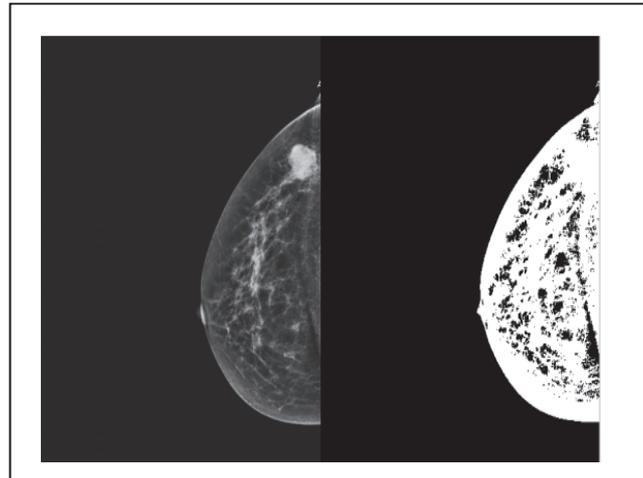


Figure 5a. filtered gray scaled image Figure 5b. binary image



Figure 6. Result of wavelet decomposition based on high frequency features.

The next step is to extract the features using wavelet transformation. The wavelet transformation extracts information features that are not clear in the raw image. Wavelet transformation can analyze the image into low and high frequency components. The needed features for classification is the high frequency components.

Two-dimensional discrete wavelet transformation is used in this paper. The output of decomposition consists of four matrices. Those are defining the approximation coefficients and detailed coefficients respectively. The detailed data matrices represent the extracted features, and those features will be the input vector for the artificial neural network system, which will be used in both training and evaluation.

When reconstructing the image based on detailed matrices of discrete wavelet transformation, the result will be much minimized and show very low quantity of high frequency data. It illustrated in figure 6.

4. Neural Network

The previous section described the preprocessing phase of the presented algorithm in this paper. The current section demonstrates the use of neural network as intelligent decision making computational engine.

This paper deployed a feed forward artificial neural network with back propagation learning algorithm. The back propagation is the most famous supervised learning algorithm. Supervised learning means that, the machine learning is done based on historical data that determines specific inputs and specific target outputs for those inputs.

A historical data was collected in this paper considering 1000 input images for different patient cases, where 700 of those are non-cancer patients and the breast undergoes different defects than this for cancer patients . A set of 300 cases was selected as cancer patients. The artificial neural network design includes designing the input vector of the network which is the detailed matrices of the high frequency decomposition of discrete wavelet

transformation that was gotten in the previous section. The output of the network is a decision that is either 1 or 0. The result of 1 means that, the image is expected to have breast cancer signs, where no cancer signs detected is reflected by the zero result.

The use of artificial neural network in supervised learning topology has many advantages with respect to the presented algorithm and the goal of this paper. This includes the complexity handling of complex input that is not easy to be processed using numerical or statistical analysis, and it is not possible or easy to be mathematically modeled. Another benefit of using artificial neural networks is to handle the non-linearity of the plant or problem model that is not easy to be handled using analytical or numerical methods of mathematics.

The selected neural network structure that was used in this paper is described in table 1. It consists of one hidden layer, one input layer, and one output layer, where the activation functions are logarithmic sigmoid, logarithmic sigmoid, and linear activation functions respectively.

Table 1. Neural network design parameters

Parameter	Value
Number of hidden layers	1
Number of neurons in first layer	240
Number of neurons in second layer	10
Used function	Logsig, logisg, purelin
Epoch	1000

5. Results

The sample data set was selected to be 1000 image related different patients. A set of 700 was selected to be non-cancer patients and a set of 300 was selected to be cancer patients. All data set was preprocessed as the early described image processing techniques and algorithm in this paper.

The output of image processing phase is the extracted features of the input raw mammographic images that are ready to be tested using neural network. A six-hundred of the non-cancer patient images set were selected for training phase, and 100 of them were left for testing purposes. Also, 2500 images of the cancer patients' images data set were used for training and the remaining 500 images were left for testing purposes.

Figure 7 illustrates the training performance graph for different learning rates (LR). Performance error is minimized with continuous epoch of training. The figure shows the effect of learning rate on the error and speed of learning. The good estimation and selection of learning rate resulting in lower error with minimal number of epochs. This means optimization in time and accuracy, where the number of epochs is related to time, and number minimal error is related to accuracy.

Figure 8 illustrates the validation test of the designed and trained neural network with different network design parameters. The training performance could be clearly understood from figure 7, figure 8, and figure 9. The training was successful in initial test and estimation of neural network performance. This was because of good design of the network as well as the rigid preprocessing that was deployed in this paper.

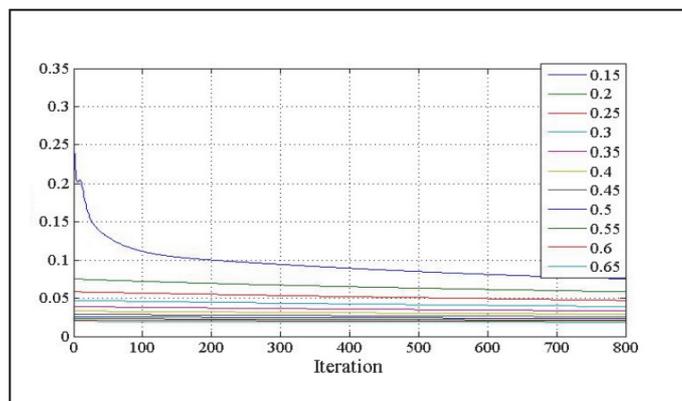


Figure 7. Training performance of the neural network versus iterations with different values of LR

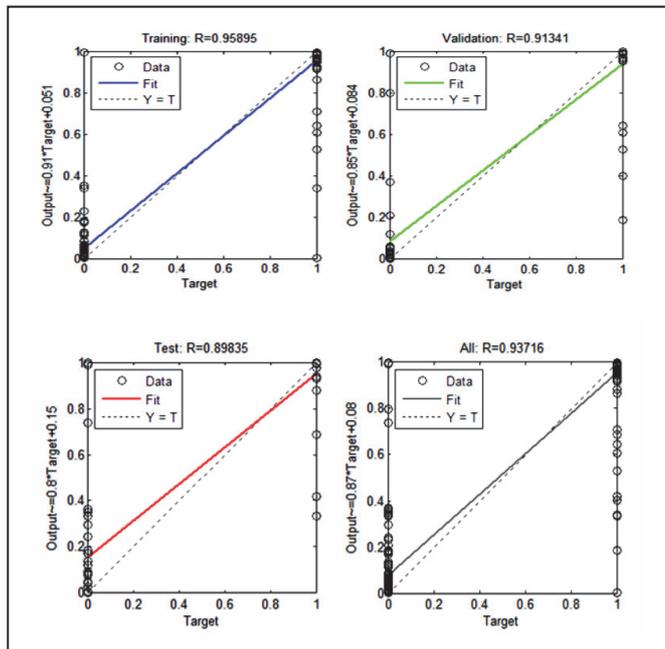


Figure 8. Validation of the resulted trained neural network

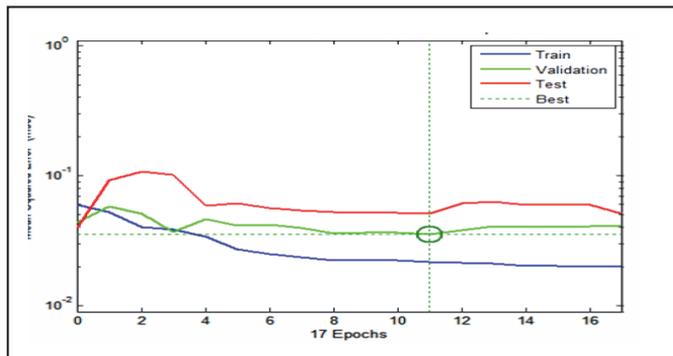


Figure 9. Accuracy of the designed neural network over testing of the dataset

Figure 9 illustrated the mean square error of testing the data set over the designed algorithm. The train data was tested separately and colored in blue. The testing data was tested separately and colored in red. And the validation result in green is the combination between both.

6. Conclusion

The complexity of decision making regarding breast cancer by the human experience of physiotherapists and radiologists makes it important to build computerized algorithms and applications that help the specialists in decision making.

The proposed algorithm of breast cancer detection gained its strength from the rigid preprocessing and feature extraction scheme as well as the use of right decision making algorithm. Morphological operations as well as good designed noise removal filter will prepare the mammographic image for feature extraction. This will put the concentration of feature extraction on ROI and meaning full data regardless of trying to extract features from raw image with random contents. The extraction of features using discrete wavelet transformation decomposition minimizes the number of features and uncovered the data that contains the information about image classification conditions.

The use of supervised training with neural networks in well-understood and designed structure helps to make the classification accuracy as high as the mean square error becomes near to ten to the power minus ten. This means that, the developed algorithm can make decisions with accuracy near to 99%. And help the therapists to save effort,

time, and human resources.

References

- Bhagyashri G. P., & Sanjeev N. J. (2014). Cancer Cells Detection Using Digital Image Processing Methods. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, 2014.
- Gayathri. B. M., Sumathi, C. P., & Santhanam, T. (2013). Breast cancer diagnosis using machine learning algorithm a survey. *International Journal of Distributed and Parallel Systems*, 4(3).
- Kalaivani, S., Prमित, C., Shikhar, J., & Rishi, G. (2017). Lung cancer detection using digital image processing and artificial neural network. *IEEE International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 2017. <https://doi.org/10.1109/ICECA.2017.8212773>
- Kother, M., Arumuga, P., & Sathik, M. (2011). Image Denoising and Enhancement Using Multiwavelet with Hard Threshold in Digital Mammographic Images. *International Arab Journal of e-Technology*, 2(1), 2011.
- Mertiz, F., Pina, R., Chou, I., & Seybold, K. (2012). Image Processing System and Methods for Aligning Skin Features for Early Skin Cancer Detection Systems. *USPTO 8194952*, May (2012).
- Ponraj, N., Jenifer, M. E., Poongodi, P., & Manharan, J. (2011). A survey on the processing Techniques of Mammogram for the Detection of Breast Cancer. *Journal of Emerging Trends in Computing and Information Sciences*, 2(12).
- Rahman, A., Nisha, K., Akter, M., & Uddin, S. (2013). Application of artificial neural network and binary logistic regression in detection of diabetes status. *Science Journal of Public Health*, 1(1), 39-43.
- Salahat, E., Saleh, H., Salahat, S., Sluzeh, A., Al-qutayri, M., Mohammad, B., & Elnaggar, M. (2016). Methods and Systems for Processing MRI Images to Detect Cancer, *USPTO 20160113546*, April (2016).
- Sheshadri, H. S., & Kandaswamy, A. (2006). Computer aided decision system for early detection of breast cancer. *Indian J Med Res.*, 124, 149-154.
- Yamamichi, Y., Umeki, M., & Akagi, E. (2009). Breast Image Processing System and Breast Image Processing Method, *USPTO 20090016580*, Jan (2009).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).