



BREAST CANCER DETECTION USING ENSEMBLE CLASSIFICATION AND EXTENDED WEIGHTED VOTING METHOD

M.Chithra Devi

Research Scholar, Department of Computer Science,
Bharathiar University,
Coimbatore, India

S.Audithan

Principal,
Sri Aravindar Engineering College,
Vilupuram, India

Abstract: Breast cancer disease increases the mortality rate of women. Early detection of cancer helps to increase the life span and decreases the mortality rate. The daubechies-8 wavelet filter, discrete wavelet transform entropies such as Shannon, Log energy and Sure entropy were applied to extract the textural features of mammogram image. The ensemble classifiers K-Nearest Neighbor, Bayes and Support Vector Machine were used to classify. The novel method of extended weighted voting method is applied to obtain the accurate result which produces the increased performance of 97.3% of results.

Keywords: Wavelets; KNN; Bayes; Support Vector Machine; Extended Weighted Voting.

I. INTRODUCTION

The medical image processing plays a vital role in medical field. To diagnose breast cancer digital mammogram images are widely preferred by radiologist. Early detection of cancer helps for treatment and reduces death rate. Women are suffered a lot compared to men by the breast cancer. Breast cancer is one type of cancer where the breast cells grow uncontrollably. To provide the maximum accuracy various algorithms were applied by various researchers. Still the experts are in meshed state to find the decision during examination of presence of cancer and its stage. In this work Daubechies-8 wavelet is used in preprocessing stage. Shannon entropy, Log Energy and Sure Entropy are applied to extract the features of given mammographic image. The classifiers such as K- nearest neighbor, Bayes and Support vector machine were processed to classify. The ensemble strategy based on novel extended weighted voting method is proposed to increase the percentage of accuracy with 97.3 %.

II. BACKGROUND WORK

In the early stages of research breast cancer detection is based on low energy X- ray mammography was in practice. Later Magnetic Resonance images, Ultra Sound images are also preferred [1]-[3]. In preprocessing stage various methods such as Binarization, Image thinning, Image gray scale extending, discrete wavelets, real valued or complex valued continuous filters, fuzzy filters were applied by many researchers [4]-[7]. Similarly, transformation techniques were used to extract physical features or textural features of an image [8]-[9]. Transformation techniques Euclidean Distance Transform, Fourier Transform, Discrete Wavelet Transform were widely used [10]-[14]. There are several classifiers to in process obtain the decision of finding the presence or absence of the cancer cells. The classification results varies based on the research work to find the whether it is benign or malignant. Some researchers provide the stages of cancer. But most of the research works was based on single classifier [15]-[19]. In general the majority voting method was also applied to find the optimized result [20]. The majority voting is the simple decision rule that selects one of many alternatives, based on the predicted classes with the most votes. In the majority

voting it does not require any parameter tuning once the individual classifiers have been trained [21]. A weighted voting method is another approach to find the optimal solution. [22]. This paper is organized as follows that provides proposed work flow, experimental results and conclusion of this research work in chapters 3, 4 and 5 respectively.

III. PROPOSED WORK

The proposed work architecture using three entropies and ensemble classifiers with extended weighted voting method for optimized result is depicted in figure 1.

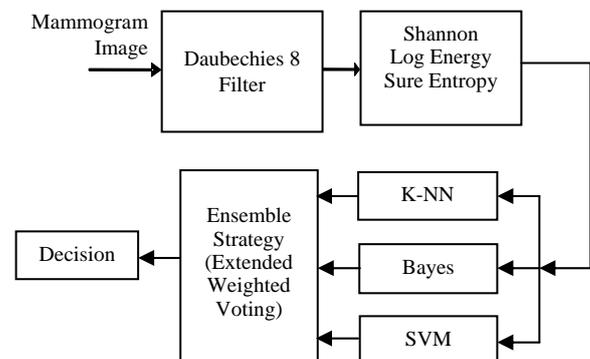


Figure 1. Proposed Architecture

A. Preprocessing

The preprocessing stage is the preliminary stage, where the given input is prepared for the next stage to compute further. The wavelet filters are widely used to clean up the image without blurring or muddling the details of a given image as input. Among various discrete wavelet filtering techniques the db8 wavelet filter is applied to get the clear image. The db8 wavelet consists of wavelet function called mother function and scaling function alias father function which generates a multiresolution analysis. In db8 the integer 8 represents the number of coefficients. The db8 wavelet supports orthonormal wavelets, thus making the discrete wavelet analysis practicable. The decomposition of the given signal by DWT is

described in terms of their basis function is given in equation 1.

$$f(x) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_n^m u_{m,n}(x) \quad (1)$$

Where, the $u_{m,n}(x) = 2^{-m/2} u(2^{-m}x - n)$ is the translations and dilation of basis function $u(x)$. The wavelet coefficients C_n^m can be computed using a pair of low and high pass filter. The output for 1 level DWT decomposition is shown in figure 2. The analysis filter bank splits a discrete image $x(m, n)$ into four sub-bands: one coarse scale (LL) and three fine scales (HL, LH, HH). If this analysis filter bank is iterated on the coarse sub-band, then the spectrum of the original image is divided by the wavelet as shown in figure 3. The subscript in figure 3 shows the scale when the sub-bands are created.

LL ₁	HL ₁
LH ₁	HH ₁

Figure 2. DWT Decomposition Level 1

LL ₄	HL ₄	HL ₃	HL ₂	HL ₁
LH ₄	HH ₄			
LH ₃		HH ₃		
LH ₂		HH ₂		
LH ₁			HH ₁	

Figure 3. DWT Decomposition Level 4

B. Feature Extraction

In this stage the required textual features are extracted. Among various textural features of an image entropy is measured to retrieve the unique data. Entropy is a measure of the number of possible arrangements in an image can have. It is the statistical measure of uncertainty or randomness that can be used to characterize the texture of the input image. The Shannon, Sure and Log entropy are applied to get the texture value of an image. It first finds the mask of an image by convert into grey scale image. Masking helps to extract out the relevant pixels in the image. The entropy measure is calculated by using the following equation.

$$E = \text{entropy}(I) \quad (2)$$

Where, E is a scalar value representing the entropy of gray scale image I.

Types of Entropy	Computation
Shannon	$-\sum_i C_i^2 \log(C_i^2)$
Log energy	$\sum_i \log(C_i^2)$
Sure entropy	$ C_i \leq \epsilon \rightarrow e(s) = \sum_i \min(C_i^2, \epsilon^2)$

Where C_i is the wavelet coefficients of a particular sub-band i with $\log(0) = 0$ and ϵ is a positive threshold value which is determined using the principle of Steins unbiased risk estimate. The estimated statistical measures of an image are passed for feature classification.

C. Feature Classification

The extracted features from the region of interest are further processed to undergo classification. The ensemble classification process is proposed in this work by applying three classifiers k-Nearest Neighbor, Bayes and Support Vector Machine. In general K-NN is a non-parametric technique widely preferred for classification. The KNN was evaluated utilizing the Euclidean distance approach. Bayesian is a probabilistic methodology to induction. It is established on the start that the amounts of interest are administered by probability distributions and that ideal choices might be made by reasoning about these probabilities together with watched information. Support vector machine (SVM) is a supervised learning algorithm which is used for effective classification. It automatically generates the hyper plane to classify the expected results of either benign or malignant. General representation for weighted voting is given in equation 2.

$$(q: w_1, w_2, w_3) \quad (3)$$

where, w_1, w_2, w_3 are weightages for classifiers. The weightage 50, 30, 20 are assigned to classifiers support vector machine, Bayes and KNN respectively. In addition to weightage the threshold value 70 is computed and the decision depends on mandatory condition assigned to each classifier.

```

Algorithm: Extended Weighted Voting (EWW)
Step 1: Start
Step 2: Get input (r1, r2, r3)
Step 3: Let w = 100
Step 4: Let w1 ← w/2
Step 5: Let w2 ← (w x 3)/10
Step 6: Let w3 ← w/5;
Step 7: Let ew = (t: w1, w2, w3)
Step 8: If ew >= 70 then
    true: If (r1 >= 50 and r2 >= 50 and r3 >= 50) then
        true: d = "Malignant"
    else if (ew > 50) then
        true: d = "Screening"
    false: d = "Negative"
    If loop ends
Step 9: Return d
Step 10: End
    
```

IV. EXPERIMENTAL RESULTS

The digitized mammographic images obtained from Mammographic Image Analysis Society (MIAS) database

Table I.Types of Entropy

were used in this work. Mammographic images have multi resolution values. The discrete wavelet transform used in this work decomposes the mammographic image into 6 levels. In general the redundant data will be created at higher resolution. In this experimental study it is found that the fourth level of decomposition of db8 filter gives better performance than other levels of decomposition. The entropy measures passed through three classifiers and the extended weighted voting method are applied to get the results. The table II depicts the accuracy level obtained at 6 different levels of decomposition using db8 filter.

Table II. Ensemble Classification Accuracy at different Levels

<i>Levels of Wavelet Decomposition</i>	<i>Db8 Classification Accuracy</i>
1	81
2	83.3
3	86.9
4	97.3
5	91.4
6	90.1

The sensitivity and specificity measures are computed. In medical domain, these two measures are mandatory to test the system performance in a medical domain. The sensitivity measure is computed for positive instances and the specificity measures for negative instances. Table 3 shows the computation of sensitivity and specificity.

Table III. Accuracy Parameters

<i>Parameter</i>	<i>Computation</i>
Sensitivity	$TP/(TP+FN)$
Specificity	$TN/(FP+TN)$

where, TP (True Positive) represents malignant mammogram is correctly classified as malignant, FP (False Positive) represents benign mammogram is incorrectly classified as malignant, TN (True Negative) is benign mammogram is correctly classified as benign. Similarly FN (False Negative) represents malignant mammogram is incorrectly classified as benign.

Table IV. Classification Accuracy based on Classifiers

<i>Classifiers</i>	<i>Accuracy at 4th Level of Decomposition</i>
K-Nearest Neighbor	76.4
Bayes	80.5
Support Vector Machine	90.2
Ensemble Classification	97.3

The region of interest is selected and passed for further feature extraction stage. The table 4 shows the experimental results obtained with individual classifier KNN, Bayes and SVM without ensemble classification. It is clear that in weighted voting, weights of voting allotted for each classifier varies based on their performance. The weight should be high for the

particular classifier among n classifiers based on the high accuracy level. The appropriate weights of votes for all the classifier are computed. Weighting problem can be viewed as an optimization problem. In addition the threshold computation is based on the mandatory condition assigned in this proposed method results better accuracy as shown in the table 4.

V. CONCLUSION

Ensemble learning applied in this work using three individual classifiers and combines their predictions, results better performance than a single classifier. Each classifier provides different contribution to the final classification result. This paper assigns priority based weights to the classifiers K-nearest Neighbor, Bayes and Support Vector Machine respectively. From the experimental result it is found at fourth level of decomposition using db8 filter and ensemble strategy based extended weighted voting method achieves better performance of 97.3 % with minimized false positive and false negative rate.

VI. REFERENCES

- [1] Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I, Stamatakis E, Cerneaz N, Kok S, Taylor P. The mammographic image analysis society digital mammogram database. In Excerpt a Medical International Congress Series 1994; 1069: 375-378.
- [2] Ibrahim N, Fujita H, Hara T, Endo T. Automated detection of clustered microcalcifications on mammograms: CAD system application to MIAS database. Phys Med Biol 1997; 42: 2577-2589.
- [3] Juan Shan, Wen Ju, Yanhui, Guoa, Ling Zhang, H.D. Cheng. 2010. Automated breast cancer detection and classification using ultrasound images: A survey, Pattern Recognition Volume 43 pp.299 – 317.
- [4] Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 1995; 90: 1200-1224.
- [5] Seng Kah Phooi, Man Zhihong, H. R. Wu. 2002. A New Approach in Fuzzy Adaptive Filtering, Fuzzy Logic Studies in Fuzziness and Soft Computing Volume 81, pp 277-287.
- [6] Taranjot Kaur. 2014. Image denoising Algorithms and DWT: A Review, Int. Journal of Computer science and Information Technologies Vol. 5 No. 5.
- [7] S.N. Yu, K.Y. Li, Y.K. Huang. 2006. Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model, Comput. Med. Imaging Graphics 30 163-173.
- [8] Eltoukhy MM, Faye I, Samir BB. A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. Comp Biol Med 2012; 42: 123-128.
- [9] Karahaliou AN, Boniatis IS, Skiadopoulos SG, Sakellaropoulos FN, Arikidis NS, Likaki EA, Panayiotakis GS, Costaridou LI. Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications. IEEE Trans Inform Technol Biomed 2008; 12: 731-738.
- [10] Mudigonda NR, Rangayyan RM, Desautels JE. Gradient and texture analysis for the classification of mammographic masses. IEEE Trans Med Imaging 2000; 19: 1032-1043.
- [11] Wei L, Yang Y, Nishikawa RM. Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. Patt Recogn 2009; 42: 1126-1132.
- [12] R.N. Strick. 1996. Wavelet transforms for detecting microcalcifications in mammograms, IEEE Trans. Med. Imaging Vol. 15 No 2 pp 218-229.

- [13] Jasmine JL, Baskaran S, Govardhan A. Non-subsampled contourlet transform based classification of microcalcification in digital mammograms. *Proc edia Engineering* 2012; 38: 622-631.
- [14] Avci D. An expert system for speaker identification using adaptive wavelet sure entropy. *Exp Sys Appl* 2009; 36: 6295-6300.
- [15] Ganesan K, Acharya UR, Chua CK, Lim CM, Abraham KT. One-class classification of mammograms using trace transform functional. *IEEE Trans Instrument Measur* 2014; 63: 304-311.
- [16] Wei L, Yang Y, Nishikawa RM, Jiang Y. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. *IEEE Trans Med Imag* 2005; 24: 371-380.
- [17] Cascio D, Fauci F, Magro R, Raso G, Bellotti R, De Carlo F, Tangaro S, De Nunzio G, Quarta M, Forni G, Lauria A. Mammogram segmentation by contour searching and mass lesions classification with neural network. *IEEE Trans Nucl Sci* 2006; 53: 2827-2833.
- [18] Mencattini A, Salmeri M, Rabottino G, Salicone S. Metrological characterization of a CADx system for the classification of breast masses in mammograms. *IEEE Trans Instrument Measur* 2010; 59: 2792-2799.
- [19] El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imaging* 2002; 21: 1552-1563.
- [20] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
- [21] L. Lam and C. Y. Suen, "Application of majority voting to pattern recognition: an analysis of its behavior and performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 553–568, 1997.
- [22] L. I. Kuncheva and J. J. Rodríguez, "A weighted voting framework for classifiers ensembles," *Knowledge and Information Systems*, vol. 38, no. 2, pp. 259–275, 2014.