selection

Pramod B. Bhalerao[1], Prof. S.V. Bonde[2]
[1]Corresponding Author, [2]Co-Author
[1]Research Scholar, Computer Science & Engg., S.G.G.S.I.E. & T, Nanded.
[2]Professor, Electronics, and Telecommunication Engg. S.G.G.S.I.E. & T, Nanded
[1,2]Vishnupuri, Nanded - 431606, (M.S.) India

Abstract: The medical image processing field is gaining more importance nowadays because of its varied variety of medical domain applications. Breast cancer is one of the common after skin cancer in the United States. Mammography is one of the techniques used. The radiologist can detect the tumor location in these mammograms, but sometimes it is difficult for them to find the tumor because of the noise. This paper focuses on feature extraction and the method of selecting features. First, ROI is extracted, and feature extraction is done with GLCM & 2D DWT followed by classification. For improving the performance, features are selected with the help of Fixed Rank Representation. Performance evaluation is done with the MIAS breast cancer database. The performance of the proposed method outperformed other methods.

Keywords: Mammography, Breast cancer database, Mammograms, tumor location

1 Introduction

breast cancer is most responsible for the mortality rate in women. Tumor detection is a tedious and challenging task because of mammogram's nature and their contrast and noise presence. Globocan work found that 1.62 million new cancer cases were observed, and in India 1, 44,937 women had cancer, and 70,000 women died, out of two women is killing due to cancer [1]. To check the presence of mass radiologist is responsible; if the radiologist misses finding the cancerous mass, it leads to false predictions. A computer-aided detection system is a valuable technique for detecting abnormal mass in mammographic images. If periodic breast screening is done, then early detection of cancerous mass may be possible, and one can even apply some measures to recover it soon. It is necessary to decrease the biopsy rate without losing malignancies to minimize patient stress and mortality. Computer-aided diagnosis (CAD) can also improve the effectiveness of mammography by aiding radiologists and its potential advantages in identifying screening mammography tumors that may otherwise be missed.

Here, the proposed work is based on using Fixed-rank representation to indirectly eliminate replicated indices from the results and increase precision. GLCM is used for feature extraction, we must choose efficient features among them, and there must be a systematic approach to selecting valuable features. Figure 1 describes operations to be performed during the overall process.

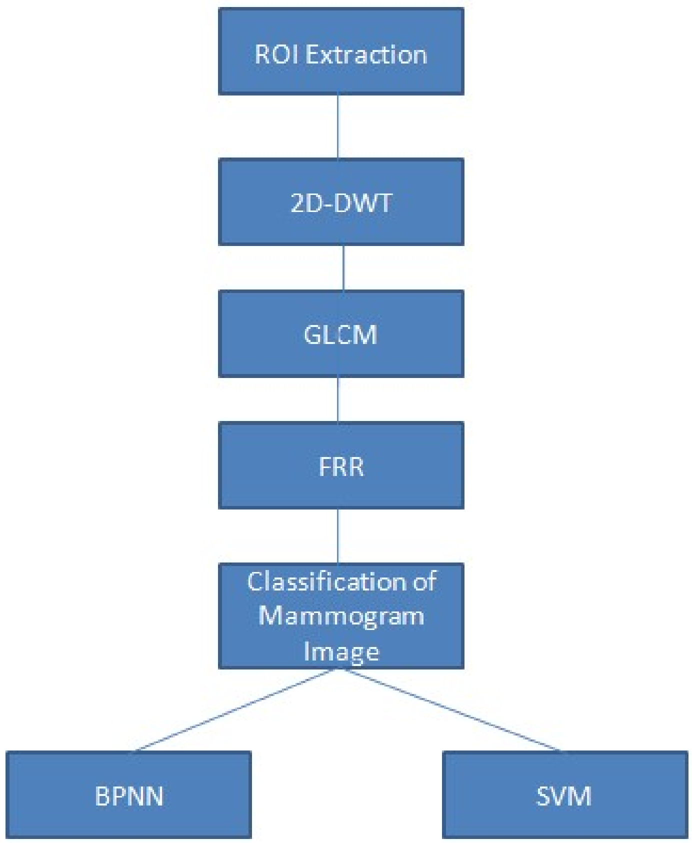


Figure 1. Working on the proposed work

In our work, starting with the input image as a mammogram, we found the ROI. Using 2D-DWT and GLCM, features are extracted, and convenient features are chosen using Fixed Rank Representation (FRR). Classifiers like the backpropagation neural network (BPNN) and the support vector machine (SVM) are used to classify masses as benign or malignant. For this experimental evaluation, MIAS [2] database is utilized.

2. Related Work

J. Anitha et al. [3] have proposed an approach; pre-processing is done to remove the noise, and segmentation is done with morphological operations; the features extracted are wavelet features, and they are compared with the GLCM features. Other authors have also done work on feature extraction, such as in [4] S. Paramkusham et al. has proposed action on the texture feature extraction and classification based on the k-means algorithm. Work done on feature selection has been given in [5-7], where each of them has used a different approach for the feature selection based on texture and geometric features or fuzzy approach. It has been found that after feature extraction, we go for the classification method in the literature [8-11] shows the work done on the classification of various ways for mammographic images. It is found that SVM gives better results than the other classifier. In [12], Shradhananda B has proposed selecting the features based on f test and t-test. In the case of extraction for the low-level feature, it doesn't involve any processing, whereas high-level features are used in [13] for examining narrow and broad weed using 2D-DWT.

3 Methodology

3.1 Preprocessing:

It has been observed that mammogram images may contain some noise, labels, or artifacts. Because of these unwanted parts, there is a problem in the identification of masses. For this cause, the Region of Interest (ROI) of these images is extracted applying some pre-processing steps, considering the estimated radius, center of abnormality. Performance analysis is done using the MIAS dataset [2]. There are 322 images in total with varied types like circumscribed, calcification, speculated, asymmetric. Every image in this is 1024 pixels x 1024 pixels. They provide ground truth details to locate the mass in the mammogram. Considering the center of abnormality X, Y Coordinates are given along with an approximate radius. With the help of these details, we can find the abnormal portion enclosing the circle, as shown in Figure. 2

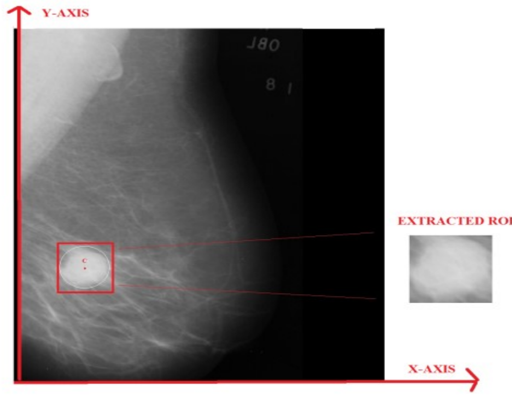


Figure 2. Extracted ROI (mdb028 MIAS)

3.2 Feature Extraction Using 2D-DWT

Image is decomposed into various sub-bands by a wavelet transformation. With the help of the wavelet transform, a thorough resolution is obtained concerning the considered small region. Every sub-band image is considered for extracting various texture features. These characteristics are used for identification and classification.

We have applied a mother wavelet function to the original function as shown below:

$$w[f(s,t)] =< f, \psi_{s,t}^{k} \geq = \int f(x)\psi_{s,t}^{k}(x)dx \quad 1)$$

This 2D-DWT is one of the decomposition techniques based on multiple levels. Every level of decomposition will create four sub- sampled parts of images, namely sub-bands. The decomposition of these sub-band images is advantageous. DWT now becomes one of the dominant tools in various applications such as video processing, image processing, and numerical analysis. one of the advantages of DWT, as compared to the discrete Fourier transform, is that it is used in performing multi-resolution analysis. With respect to the frequency and time domain. The image consists of a 2-D signal, namely a column and a row, and it is decomposed using 2D DWT. Convolution is done on filters for gaining 2-D transformation.

The 2-D DWT offers four sub-band images in each step of decomposition. . LL refers to image approximation. LH corresponding to the horizontal details contains high vertical frequency, low horizontal frequency image information. HL consists of low vertical frequency and high horizontal frequency, corresponding to the vertical information. HH contains high vertical and high horizontal frequencies, corresponding to diagonal data. After that, successively LL sub-band image decomposed using 2-D DWT to get another decomposition level; figure 3 denotes the decomposition up to level 2.

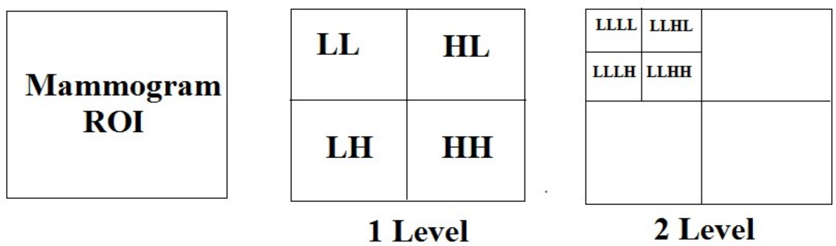


Figure 3.Decomposition of Image using 2D-DWT up to level 2

3.3 Gray level Co-occurrence Matrix

This method is utilized for extracting the features related to texture. The GLCM technique is a method of statistical extraction of features. Features are important pieces of information. The processing time and complexity will be reduced by removing the features. For extracting features, the spatial relationship of the pixel is taken into consideration [14].

The image matrix of the co-occurrence is calculated by using GLCM. The relationship of pixels with other pixels can be determined in terms of distance and angle. The formula for GLCM of an image with size $N \times N$ is given by

p(i,j) = $\sum_{i=1}^{N}\sum_{j=1}^{N} 1, I(x,y) = i \text{ and } I(x+\Delta_x, y+\Delta_y$ 2)

o, otherwise

p(i, j) signify the joint probability of occurrence at an offset (Δx, Δ y) with spatial positions x & y and having intensities i & j. The distance d and the angle θ between the pixels to its neighborhood are specified by the offset (Δx, Δy).GLCM performance is better then as well as linear discriminant analysis principal component analysis[15]. The amount of different gray levels of the quantized image is represented by G. P_i(i) and P_i(j) represent the marginal probability matrix. The following notations represent the various features.

P_i(i) = $\sum_{i=1}^{G} p(i,j)$ 3)

P_i(j) = $\sum_{i=1}^{G} p(i,j)$ 4)

3.4 Feature Selection and Classification

After the features are selected, we found that there are so many repeated indices are there we can remove and improve the accuracy with the help of FRR (fixed Rank Representation). Followed by feature collection, these characteristics are passed on to classifiers for classification. Two classifiers, backpropagation neural network (BPNN) and support vector machine (SVM), are used to classify the tumor type. The evaluation shows that SVM gives better performance.

3.5 Fixed Rank Representation

In the case of unsupervised learning for image data, there is a problem with subspace clustering. There is currently a new technique based on sparsity, such as sparse subspace clustering (SSC). Subspace clustering is one of the methods in unsupervised learning similar to feature extraction. Now a day there is advancement in ranking and sparsity. But these techniques are more expensive. Low-rank representation (LRR) is one of the methods proposed based on sparse subspace clustering. LRR has insufficient data sampling. A similar kind of problem is also found in other forms. Some of the techniques are presented; among them, factorization is one way used in Fixed rank representation (FRR).

This FRR specifically represents products of low-rank matrices. So FRR solution is more optimal than LRR; thereby, we can solve data sampling. FRR even avoids some complex computations in LRR. In computer vision and image processing, subspace divisions are essential tasks to segment the required information and recover the subspaces. Fixed rank representation is one of the subspace division approaches to solve ineffectively inspecting in LRR. In most of the applications related to segmentations, it achieves good results than that of LRR. Thus the measure problem of insufficient sampling issue is solved by FRR [16]. It uses mainly a variable of two matrices for the approximating coefficient of a low-rank.

3.5.1 The Basic FRR Model

As of LRR, Z's value we have to minimize to diminish the Frobenius norm rather than the nuclear norm. Optimize Z as

$$\min_{z,z'} ||z - z'||^2_{F'}, s.t. X = XZ . rank\ (z') = m \qquad 5)$$

It is also represented as a matrix of product for $Z' = LR$, Here $L \in R^{n}$and $R \in R^{n}$. If we replace Z' by using LR then we come to we turn up towards FRR model

$$\min_{Z,Z',R} ||Z - LR||_{F'}^{2}, s.t. X = XZ \qquad 6)$$

3.5.2 Sparse Regularization

in most applications, data will get corrupt because of a little bit of noise or some outliers. We model this as a new phrase as E. Consider the optimized equation as

$$\min_{Z,L,R,E,} ||Z - LR||_{F'}^{2} + \mu||E||, \qquad 7)$$

$$s.t. X = XZ + E, 1_{n}^{T}Z = 1_{n}^{T}$$

Considering the value of $\mu > 0$ for balancing the effect of two different expressions. Here we approve the $l_{2,1}$ norm for characterizing corruption because it can fruitfully recognize the indices in outliers also remove little noises [17]. Algorithm 1 summarizes the whole FRR based subspace clustering framework.

Algorithm A) For Fixed Rank Representation in subspace clustering

Input: Consider $X \in R^{d}$ from k subspaces, some of the data points are sampled.

Following steps we follow

1: work out (7) to get L^{*}, R, Z^{*}

2: a graph is build using $(|Z^{*}| + |Z^{*}|)'$ OR $(|L^{*}R^{*}| + (L^{*}R^{*})'$ A similarity matrix.

3: For this graph, relate NCut to get clustering.

We can utilize FRR for feature extraction as an extension.

Algorithm B)

- **Algorithm 2 Evaluating (7) using ADM-type**

Input: scrutiny matrix $X \in R^{dxn}$, $\in_{1}, \in_{2} > 0$, $m > 0$, $parameters\ \beta > 0\ and\ \rho$

Initialization: Initialize

$$Z_{0} \in R^{nxn}, L_{0 \in} R^{nxm}, R_{0} \in R^{mxn}, E_{0} \in R^{dxn}, \Lambda_{0} \in R^{dxn}\ and\ \Pi_{0} \in \mathrm{R}^{1}$$

While: until converged **do**

1: Revise (R, Λ, Z, L, E, Π) using equation 8) to 13)

2: ensure the conditions of convergence:

$$||X - XZ_{+} - E_{+}||_{\infty} \leq and\ ||1_{n}^{T}Z - 1_{n}^{T}||_{\infty} \leq \in_{2}$$

end of while

Output: L^*, Z^*, R^*

here,

$$L \leftarrow ZR = ZR^T(RR^T) \quad 8)$$

$$R \leftarrow LZ = (L^T L)L^T Z \quad 9)$$

$$Z \leftarrow (2I_n + \beta(X^T X + 1_n 1_n^T)^{-1}B \quad 10)$$

$$E \leftarrow arg_E \min \mu\|E\| + \frac{\beta}{2}\|C - E\|_F^2 \quad 11)$$

$$\Lambda \leftarrow \Lambda + \beta(X - XZ - E) \quad 12)$$

$$\Pi \leftarrow \Pi + \beta(1_n^T Z - 1_n^T) \quad 13)$$

4 Results and Discussion

We have performed a set of steps in experimental evaluation; Fig 1 represents the different steps involved in the process

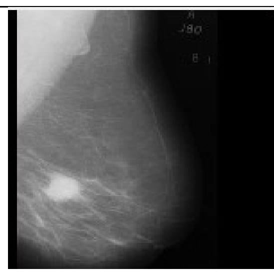
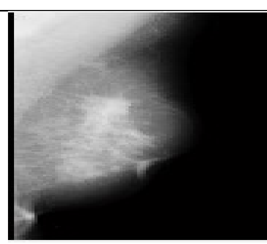
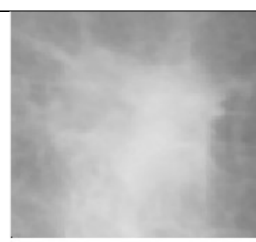
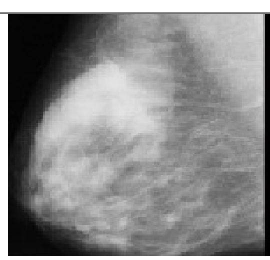
Image	Input Image	Processed (ROI)
mdb028 (MIAS)		
mdb178 (MIAS)		
mdb115 (MIAS)		

Fig 4. The input image and processed ROI for the MIAS dataset image.

Table 1.Performance comparison with state-of-art methods.

	the t-test (BPNN)	f-test (BPNN)	Only GLCM (BPNN)	GLCM with FRR (BPNN)	GLCM with SVM	GLCM with FRR and (SVM)
Accuracy	88.2	82.4	57.02	66.31	86.18	89.4
Specificity	88.9	77.8	55.63	59.31	85.77	89.2
Sensitivity	87.5	87.5	55.63	59.31	85.77	89.1

Based on the values, we get Fig 3. Represents the graphs we get.

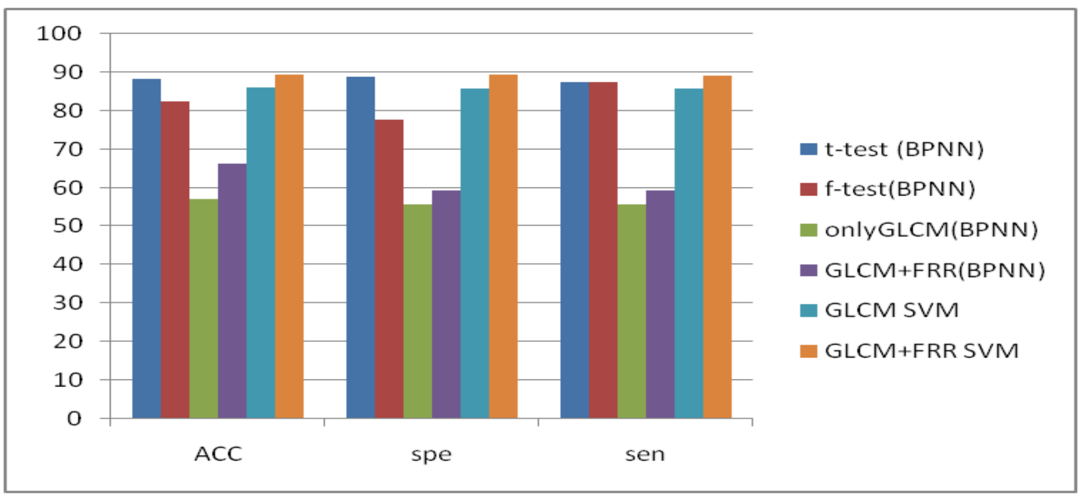


Fig 5. Representing the results in a statistical format

5. Conclusion

Medical image segmentation and detection of masses in mammograms still need some improvement. In our present work, we gave a hybrid approach using fixed rank representation for feature selection. Initially, ROI is extracted, and the Feature matrix is calculated based on GLCM and 2D-DWT; for reducing the number of features, fixed rank representation is used for feature selection. Experimental results are given based on Accuracy, Sensitivity, and specificity. It has been found from results that our proposed approach based on GLCM and FRR gives better classification accuracy (89.40%) for SVM as compare to the BPNN classifier(66.31%). Similarly, other performance parameters provide better results for the presented approach.

REFERENCES

1. "International Agency for Research on Cancer (iarc),Globocan Project 2012, and World Health Organization: cancer fact sheets".
2. http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html
3. Anitha, J., & Dinesh Peter, J. (2012). A wavelet-based morphological mass detection and classification in mammograms. 2012 International Conference on Machine Vision and Image Processing (MVIP).
4. S. Paramkusham, K. M. M. Rao and B. V. V. S. N. P. Rao, "Novel technique for the detection of abnormalities in Mammograms using texture and geometric features," 2015 (ICMOCE), Bhubaneswar, 2015, pp. 150-153.
5. K. T. Raja Keerthana and K. Thangavel, "Feature selection in mammogram image using a rough set approach," 2011 National Conference on Innovations in Emerging Technology, Erode, Tamilnadu, 2011, pp. 147-152.
6. A. Heshmati, R. Amjadifard, and J. Shanbehzadeh, "ReliefF-Based Feature Selection for Automatic Tumor Classification of Mammogram Images," 2011 7th Iranian Conference on Machine Vision and Image Processing, Tehran, 2011, pp. 1-5.
7. R. B. Dubey, M. Hanmandlu and S. Vasikarla, "Features Selection in Mammograms Using Fuzzy Approach," 2014 11th International Conference on Information Technology: New Generations, Las Vegas, NV, 2014, pp. 123-128.
8. C. Juarez, M. E. Castillo, and V. Ponomaryov, "Classification of Masses in Mammography Images Using Wavelet Transform and Neural Networks," 2007 International Kharkov Symposium Physics and Engrg. of Millimeter and Sub-Millimeter Waves (MSMW), Kharkov, 2007, pp. 956-958.
9. N. Alpaslan, A. Kara, B. Zencir and D. Hanbay, "Classification of breast masses in mammogram images using KNN," 2015 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, 2015, pp. 1469-1472.

10. D. Tralic, J. Bozek and S. Grgic, "Shape analysis and classification of masses in mammographic images using neural networks," 2011 18th International Conference on Systems, Signals and Image Processing, Sarajevo, 2011, pp. 1-5.
11. N. Alpaslan, A. Kara, B. Zencir and D. Hanbay, "Classification of breast masses in mammogram images using KNN," *2015 23rd* Signal Processing and Communications Applications Conference (SIU), Malatya, 2015, pp. 1469-1472.
12. Shradhananda B, B Majhi," Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer", Neurocomputing154(2015)1–14
13. K. H. Ghazali, M. F. Mansor, M. M. Mustafa, and A. Hussain, "Feature Extraction Technique using Discrete Wavelet Transform for Image Classification," *2007* 5th Student Conference on Research and Development, Selangor, Malaysia, 2007, pp. 1-4,
14. A. E. Minarno, Y. Munarko, A. Kurniawardhani, F. Bimantoro, and N. Suciati, "Texture feature extraction using co-occurrence matrices of sub-band image for batik image classification," 2014 2nd International Conference on Information and Communication Technology (ICoICT), Bandung, 2014, pp. 249-254.
15. Eleyan, Alaa&Demirel, Hasan. (2011). Co-occurrence matrix and its statistical features as a new approach to face recognition. Turkish Journal of Electrical Engineering and Computer Sciences. 19. 97-107. 10.3906/elk-0906-27.
16. R. Liu, Z. Lin, F.D.l. Torre, Z. Su, Fixed-Rank Representation for Unsupervised Visual Learning, in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'12, Providence, RI, USA, 2012, pp. 598-605.
17. G. Liu, H. Xu, and S. Yan. Exact subspace segmentation and outlier detection by low-rank representation. submitted to JMLR (arXiv:1109.1646), 2011.