International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-1, May 2020

Segmentation and Classification of Mammogram into Normal and Abnormal using Texture Features

B. V. Divyashree, C. A. Soujanya, M. R. Keerthana, M. Naveen, G. Hemantha Kumar

Abstract: Breast cancer is known to be a fatal disease since decades in women worldwide. Mammography is an effective tool used for the detection of breast cancer in the early stage. Computer aided tools helps medical field by ruling out the false identification of cancer cells in mammograms. Breast region extraction and classification of the extracted region into normal and abnormal is a crucial step in mammographic based diagnosis of breast cancer. Hence, in the proposed paper a method for segmentation of breast region and classification of breast region is presented. Breast region extraction is performed using Otsu's thresholding method and intensity adjustments, enhancement is performed by Contrast Limited Adaptive Histogram Equalization (CLAHE). Gray Level Co-Occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP) features are extracted to classify the breast region using K-Nearest Neighbors (KNN) classifier. The proposed algorithm is tested on Mammographic Image Analysis Society (MIAS) dataset, obtained minimum Root Mean Square Error (RMSE) and maximum Peak Signal-to-Noise Ratio (PSNR). For classification, 80.12% of accuracy is obtained with TPR and FPR of about 0.8317 and 0.3412 respectively.

Keywords : Mammogram, Segmentation, Feature extraction, Otsu's thresholding, Gray Level Co-Occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP).

I. INTRODUCTION

Cancer is one amongst the many mortal diseases since decades. Breast cancer is one of the major cancer types in women around the world [1]. Breast cancer is completely curable if detected in the preliminary stages. Mammography is found to be a successful radiographic screening tool for timely diagnosis of breast cancer [2]. A computer-aided diagnosis system helps in identification of early signs of breast cancer. It also assists the radiologists to detect a large number of mammograms at once. It is difficult to explore the hidden characteristics of mammograms as it appears in low contrast. So, preprocessing and classification of the mammograms are vital steps that are required in automatic breast cancer detection [3].

In the-state-of-the-art preprocessing techniques have been reported for mammographic images as it influences the accurate detection of cancer [4, 5]. Various enhancement techniques including spatial and frequency domain techniques were adopted in the literatures [6]. A comparative study on wave-let based enhancement and morphological operators were presented in [7]. Contrast enhancement functions and optimal adaptive neighborhood functions were adopted. But, the method failed in enhancing the barely seen features in an image [8]. Algorithms for background suppression and enhancement were proposed using dual-tree complex wavelet transform. It eliminates the limitation faced by linear filtering technique [9, 10]. Classification of mammographic image into normal and abnormal mammograms masses was presented by using SVM, fuzzy logic, neural network and genetic [11, 12]. Existing studies focuses on classification of mass but, in the proposed paper, classification of breast region into normal and abnormal region is performed as different abnormalities appear in various forms in breast region.

In the proposed paper, artifacts are removed by suppressing the background, pectoral muscles are removed and obtained breast region. Enhancement and classification of estimated breast boundary are performed. Background suppression is accomplished using Thresholding operation, pectoral muscle removal by Otsu's Thresholding and intensity adjustments, enhancement of breast region by Contrast Limited Adaptive Histogram Equalization (CLAHE), feature extraction is performed using Gray Level Co-Occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP). K-Nearest Neighbors (KNN) classifiers are used to classify the breast region of mammograms into normal and abnormal.

II. PROPOSED METHODOLOGY

Mammographic diagnosis of cancer demands proper and efficient image analysis, correct analysis of clinical evidences and classification of different tissues which is an erroneous task for radiologists and physicians [13].

Manuscript received on April 02, 2020.
Revised Manuscript received on April 15, 2020.
Manuscript published on May 30, 2020.

* Corresponding author: divyashreenivas@gmail.com

B. V. Divyashree, DOS in Computer Science, University of Mysore, Manasagangothri, Mysuru – 570006, Karnataka, India.

C. A. Soujanya, DOS in Computer Science, University of Mysore, Manasagangothri, Mysuru – 570006, Karnataka, India.

M. R. Keerthana, DOS in Computer Science, University of Mysore, Manasagangothri, Mysuru – 570006, Karnataka, India.

M. Naveen, DOS in Computer Science, University of Mysore, Manasagangothri, Mysuru – 570006, Karnataka, India.

G. Hemantha Kumar, DOS in Computer Science, University of Mysore, Manasagangothri, Mysuru – 570006, Karnataka, India.

CAD based techniques helps for mammographic analysis and accurate breast cancer detection in the early stage [14].

The proposed method targets to segment the region of interest (breast region), enhancement of breast region, feature extraction and classification of the region of interest into normal and cancer as shown in figure1.

In this work, Mammographic Image Analysis Society (MIAS) database are considered for the experimentation [15]. MIAS is publicly available dataset where the images are digitized and represented to 50 micron pixel and 8-bit word of each pixel respectively. It contains 322 mammographic images in total, out of which 56 are categorized as malignant, 202 as normal and 64 as benign.

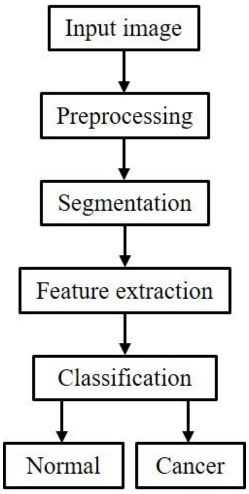


Fig 1: Overall block diagram of the proposed work

A. Preprocessing and Segmentation

Mammographic image contains various artifacts and labels that are having similar intensities as tissues in the breast region. Hence to obtain the accurate breast boundary, artifacts are removed using thresholding method. It divides the image into foreground (breast region + pectoral muscle) and background regions.

$$I(x,y) = \begin{cases} 1, \text{ if } I(x,y) > T \\ 0, \text{ otherwise} \end{cases} \quad (1)$$

For final refinement of the foreground regions, area opening and morphological closing operation is performed. Foreground region obtained from the previous step contains both breast region and pectoral region. Pectoral muscle regions are similar to tissues in the breast region. Hence to obtain only breast region (ROI) from the foreground region, pectoral muscle regions are removed. Since pectoral muscle appear in different shape, size and orientation, suppressing it is a bit challenging task. The input image is binarized to predefined threshold level to normalize the luminance of the pixel [B1]. Median filter applied to the original background image eliminated noises and highlighted the sharpness of it. Later, contrast adjustment performed twice changes the intensity values to the different values. In the next step, Otsu's thresholding method is adopted to divide the image into two regions (pectoral region and dense tissues inside breast region). Otsu's thresholding method chooses threshold value which is the minimum and maximum values of input data based on the histogram of the image and produced two regions [B2]. Finally, difference between image B1 and image B2 resulted with breast region. To remove discontinuities in the obtained image morphological closing and holes filling operations are inculcated. Fig 2 depicts the results of segmentation.

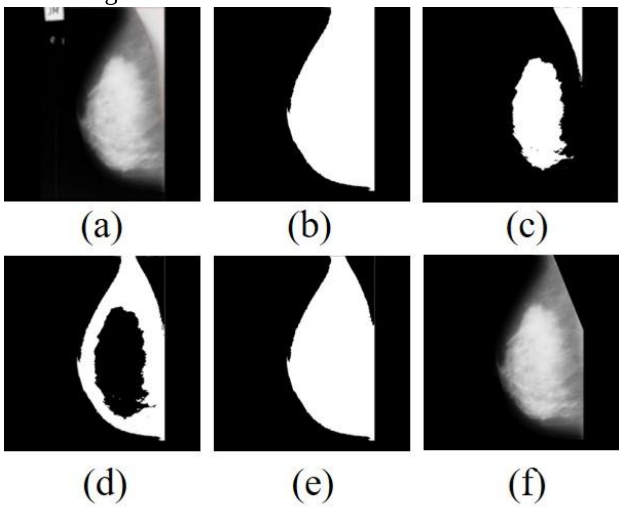
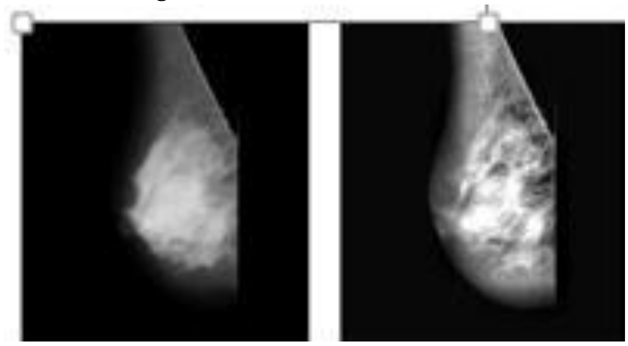


Fig 2: (a) Original image with ground truth, (b) Binary image, (c) Otsu's thresholded image, (d) Subtracted image, (e) Holes filled image, (f) Segmented breast region.

The segmented breast region is enhanced using CLAHE where the histogram is terminated at some threshold and equalized. CLAHE helps in limiting the contrast by reducing the noise amplification in mammogram. The enhanced image is as shown in Fig 3.



(a) (b)

Fig 3: (a) Original image (b) Enhanced image

B. Feature extraction

Radiologist depict normal and abnormal mammogram by their texture properties and gray levels. Abnormal tissues may reside in any part of mammogram with different textures. Hence in the presented work, we concentrated on extracting features from breast region using GLCM, LBP and HOG.

a. GLCM

GLCM features are most commonly used methods for texture analysis. GLCM is used to compute statistical features based on gray intensity levels. It analysis the texture features based on different gray level combinations [16].

Retrieval Number: A1964059120/2020©BEIESP
DOI:10.35940/ijrte.A1964.059120
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In the proposed paper, under GLCM we have extracted five statistical features namely energy, homogeneity, correlation, contrast
and energy from enhanced breast region of mammographic image. The equations of the GLCM features are listed below.

$$Energy = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2 \tag{2}$$

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \tag{3}$$

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \tag{4}$$

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \tag{5}$$

$$Energy = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2 \tag{6}$$

b. HOG

The Histogram of Oriented Gradients (HOG) is commonly used feature descriptor to identify object. It helps in counting the incidences of gradient orientation in confined area of image. The method is analogous to edge orientation histograms, shape context and Scale-Invariant Feature Transform (SIFT) descriptor [16]. The modification is that HOG is computed on consistently spread out dense grid cells and produces feature vector. For improving the accuracy, it uses local overlapping contrast normalization. In the proposed method, HOG features extracted from the segmented and enhanced breast region provided local shape information from the breast region.

c. LBP

LBP feature is most widely used feature for texture analysis especially for recognition purposes. The extracted features are useful for classifying the mammographic image into normal or abnormal mammogram. In the proposed work, given a pixel in the input image, an LBP code is calculated by its neighbor pixels considering P (number of neighbors) and R (Radius of comparison) parameters. It uses each pixel as a threshold and transforms its neighborhood (3x3) into 8-bit binary code. The static order of this binary code helps in retaining the texture information around the pixels. The feature vectors obtained after extracting features using GLCM, HOG and LBP feature descriptors, classifiers are applied to classify the mammographic image.

C. Classification

After getting the features from the breast region of the mammographic image, classification is performed to classify into normal or abnormal breast region. Classification of an image in the presented paper is a discriminating process of classifying mammographic breast region into normal or abnormal. KNN classifier performs better compared to the various other classification algorithms for the proposed work. KNN classifier is well known simplest supervised machine learning algorithms used for classification [17]. Feature vectors and labels are stored in the training phase and classified the objects based on the labels of its K-Nearest Neighbors. It works based on feature similarity and classifies new data point.

III. EXPERIMENTAL RESULTS

For validation of image enhancement algorithm Peak Signal-to-Noise Ratio (PSNR) and Root Mean Square Error (RMSE) parameters are considered.

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j) - y(i,j))^2}$$

$$PSNR = 10 \log \frac{(2^n-1)^2}{RMSE}$$

The experimental results tabulated in Table 1 shows the comparison between PSNR and RMSE values obtained for enhanced breast region.

The results tabulated depicts the minimum RMSE value and maximum PSNR value that helps in better discrimination of the features in the breast region.

Later, the classification accuracy is measured by the labels present in confusion matrix as shown in Table 2.

Table 1: PSNR and RMSE value of ten images of MINI-MIAS database.

Images from database	PSNR	RMSE
mdb001	26.0345067630773	1.7795
mdb002	25.7167397376981	1.2557
mdb003	25.9181817458174	1.9517
mdb004	25.8732850106097	1.0189
mdb005	25.7822783960680	1.1560
mdb006	25.7995628476717	1.1299
mdb007	25.9353994655914	1.2925
mdb008	25.8422870538396	1.0652
mdb009	25.8239076602645	1.0766
mdb010	26.0475439540422	1.7603

Table 2: Confusion matrix

Actual /predicted	Normal	Abnormal
Normal	TP	FP
Abnormal	FN	TN

Among 322 images in the dataset, 202 images are listed as normal and 120 images are listed as abnormal.

$$TPR = \frac{TP}{(TP + FN)}$$

$$FPR = \frac{FP}{(FP + TN)}$$

$$ACC = \frac{TP + TN}{TP+FP+TN+FN}$$

$$TNR = \frac{TN}{(TN + FP)}$$

$$FNR = \frac{FN}{(FN + TP)}$$

Accuracy, TPR, FPR, FNR and TNR parameters are tabulated in Table 3. The equations of the parameters are listed above.

Table 3: Confusion matrix

Actual /predicted	Normal	Abnormal	Total
Normal	173	29	202
Abnormal	35	85	120
Total	208	114	322

Table 4: Overall TPR, FPR, FNR and TNR rates

Accuracy (%)	TPR	FPR	TNR	FNR
80.12	0.8317	0.3412	0.7456	0.1682

The experimental results yeided accuracy of about 80.12% with TPR=0.8317 and FPR=0.3412. The cases that produced less classification result is because of the extreme noises present in the dense and extremely dense breast density types.

IV. DISCUSSION AND CONCLUSION

Breast cancer is mortal disease in women across the globe, if treatment is not given in time. Nevertheless, early detection and treatment increases the survival rate. Preprocessing, segmentation and classification are the important steps in medical diagnosis of mammograms. Accurate preprocessing, segmentation, feature extraction and classification can be expected by the adoption of proper techniques [18]. In this paper, thresholding method and Otsu's thresholding methods are adopted for segmentation, enhancement is performed using CLAHE. GLCM, HOG and LBP features are extracted from segmented region, classified into normal and abnormal breast region using KNN classifier. This paper focuses mainly on the feature extraction and classification of breast region into normal and abnormal that helps the radiologist in classifying the mammogram into normal and abnormal mammogram. Overall accuracy obtained is about 80.12% with TPR=0.8317 and FPR=0.3412. This benefits the medical field by increasing the survival rate of a patient by early diagnosis. In future better features extraction technique can be adopted for improving the classification accuracy and can also focus on density classification.

ACKNOWLEDGMENT

The B.V. Divyashree first author would like to thank the Ministry of Tribal Affairs, Govt. of India, for awarding the National Fellowship (201718-NFST-KAR-00159) to carry out this research work.

REFERENCES

1. DeSantis, C., Ma J., Bryan L., Jemal A.: Breast cancer statistics, 2013, CA A cancer journal for clinicians, 64(1), 52-62 (2014).
2. Oeffinger, K.C., Fontham, E.T.H., Etzioni, R.: Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA, 314(15), 1599-1614 (2015).
3. Moghbel, M., Chlayee, Ismail, N., Yuan wen hau.: A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection /diagnossi of breast mammography (2019).
4. Ferrari RJ, Rangayyan RM, Desautels J E, Borges RA, Frere AF, Automatic identification of the pectoral muscle in mammograms, IEEE Transactions on Medical Imaging 2004;23(2), pp.232-245.
5. Liu CC, Tsai CY, Liu J, Yu CY, Yu SS. A pectoral muscle segmentation algorithm for digital mammograms using Otsu thresholding and multiple regression analysis. Computer and Mathematics with Applications 2012;64(5), pp.100-1107.
6. Yashwanthi Sivakumari, P. Sudharsan, " Comparison of Diverse En-hancement Techniques for Breast Mammograms", International Journal of Advance Research in Computer Science and Management Studies, Volume 1, Issue 7, pp. 400-407, December 2013.
7. Subodh Srivastava, Neeraj Sharma, S. K. Singh, and R. Srivastava, "A com-bined approach for the enhancement and segmentation of mammograms us-ing modified fuzzy C-means method in wavelet domain," J Med Phys. Vol. 39(3): pp. 169–183, 2014.
8. Dhawan A. P., G. Buellon, and R. Gordon, "Enhancement of mammographic feature by optimal adaptive neighbourhood image processing," IEEE Trans. Med. Imag., Vol. MI-6, No. 1, pp. 82–83, 1986.
9. Tomklav Stoji C, Irini Reljin, Branimir Reljin, "Local contrast enhancement in digital mammography by using mathematical morphology," IEEE Transactions, 2005.
10. Bhattacharya, Debmalya, Mrs Jibanpriya Devi, and Ms Payal Bhattacherjee. "Brain Image Segmentation Technique Using Gabor filter parameter.", American Journal of Engineering Research (AJER) Vol-02, Issue-09, pp.127-132, 2009.
11. Rajkumar, T. M. P., and Mrityunjaya V. Latte. "Adaptive Thresholding Based Medical Image Compression Technique Using Haar Wavelet Based Listless SPECK Encoder and Artificial Neural Network." Journal of Medical Imaging and Health Informatics 5, no. 2 (2015): 223-234.
12. Cheng H. D., J. Shan, W. Ju, Y. Guo and L. Zhang, "Automated breast cancer detection andclassification using ultrasound images: A survey," Pattern Recognition, vol. 43, no. 1, p. 299-317, 2010.
13. Pawar, M.M., Talbar, S.N.,Genetic Fuzzy System (GFS) based Wavelet Co-occurrence Feature selection in ammogram Classification for Breast Cancer Diagnosis, Perspectives in Science (2016),http://dx.doi.org/10.1016/j.pisc.2016.04.042.
14. Burhenne L, Wood S, D'Orsi C, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology 2000;215(2):554–562. [PubMed:10796939].
15. Suckling, J. Parker, D.R. Dance, S. Astley, I. Hutt, C.R.M. Boggis, I. Ricketts, E. Stamatakis, N. Cernaez, S.L. Kok, P.Taylor, D. Betal, J. avage, The mammographic image analysis society digital mammogram database,in: Proceedings of the 2nd International Workshop on Digital Mammography, York, England, 10-12 July 1994, Elsevier Science, Amsterdam, 1994, pp. 375– 378.
16. Tatikonda, K. C., Bhuma, C. M., & Samayamantula, S. K. (2018). The Analysis of Digital Mammograms Using HOG and GLCM Features. 2018 9th International Conference on Computing, Communication and Networking Technologies(ICCCNT).doi:10.1109/icccnt.2018.8493809.
17. Onel Harrison. (2018). Machine Learning Basics with the K-Nearest Neighbors Algorithm. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761.
18. Divyashree, B.V., Amarnath, R., Naveen, M., Hemantha Kumar, G. (2018). Novel approach to locate region of interest in mammograms for Breast cancer. International Journal of Intelligent Systems and Applications in Engineering, *6*(3), 185-190.