Breast Tissue Characterization using Combined K-NN Classifier

K. Vaidehi^{1*} and T. S. Subashini²

¹Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, 608002, India; vainakrishna@gmail.com ²Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, 608002, India

Abstract

Worldwide, breast cancer is one of the top two lethal diseases among women. Breast tissue density is the important risk indicator of breast cancer. Digital Mammography technique is used to detect the breast cancer at its benign stage. Computer Aided Diagnosis (CAD) tools aids the radiologist for an accurate diagnosis and interpretation. In this work, Statistical features are extracted from the Region of Interest (ROI) of the breast parenchymal region. K-NN with three different distance metrics namely Euclidean, Cosine, City-block and its combination is used for classification. The extracted features are fed into the classifier to classify the ROI into any of three breast tissue classes such as dense, fatty, glandular. The classification accuracy obtained for combined k-NN is 91.16%.

Keywords: Breast Density, K-NN, Mammography, Statistical Descriptors

1. Introduction

According to the International Agency for Research on Cancer (IARC) statistics, 8.2 million cancer-related deaths occurred in 2012. In India, number of breast cancer cases will be estimated to double in 2025¹. Early stage of detection is the only way to prevent and protect us from breast cancer. Among the different techniques for identifying breast cancer, digital Mammography is the most widely used screening tool. CAD technique could act as a second reader for assisting radiologist to identify the abnormalities in the mammogram.

Mammogram, the X-ray image of the breast is the most widely used screening technique for early breast cancer detection. The different breast tissue namely dense, glandular and fatty is X-rayed differently due to the fact that fatty breast allow X-rays to permeate things it forming dark areas on a mammogram which allows better lesion detection. Mammogram determines the degree of density in the breast image. Fatty tissue is black in color and fibro glandular tissue is white on the mammogram². Therefore intensity based statistical features are useful for measuring the mammogram intensity variation. In this study, statistical features such as mean, standard deviation, skewness and kurtosis are used for feature extraction. K-NN with single distance may not sufficient to calculate the accurate result since K-NN with three different distance measures are used and the majority between the two different distances are considered as a final result.

2. Background

Many Computer aided diagnosis techniques for breast cancer are discussed in³ and this work vividly reviewed about enhancement, segmentation, detection and identification of suspicious regions of the breast. Two commercially available CAD systems R2 ImageChecker (version 8.3.17) and iCAD Second Look (version 7.2-H) are compared in⁴ and it concluded no major differences between the two systems. Spatial gray level dependency matrices were constructed and descriptors are extracted from it to classify the breast tissue⁵. SFS with k-NN and C4.5 classifiers are using the morphological and texture features to classify the breast tissue based on BIRADS category is carried out in⁶. The graph cut technique is proposed for visualizing the breast anatomical regions in⁷.

The work in⁸ computes six statistical features and classifies the breast tissue based on the histogram and obtained 80% classification accuracy. Histogram and accumulative histograms were utilized to estimate the breast density based on gray scale statistics in⁹. Statistical technique to segment the mammograms based on the breast density is applied in¹⁰, where Karhunen Loeve based model and linear discriminant model was employed to classify breast density, taking neighborhood pixels into account. They obtained better results with the principal component analysis model. Fractal related features and SVM classifier is employed for characterization of breast density in¹¹. The work in12 utilizes morphological and texture descriptors and sequential forward selection classifier for classification. Scale invariant feature transform, local binary patterns and texton histograms are extracted and modeled with SVM classifier for breast density classification¹³. In¹⁴ features were calculated from the statistical measures and features are classified using SVM.

3. Method

The various phases of the proposed method are shown in Figure 1.

This method has been tested on the Mini-Mias database mammograms¹⁵. Since the upper portion which contains the predominant density pectoral muscle may hinder the classification process, the ROI is chosen for the lower portion of the mammogram which is devoid of pectoral muscle and other background region.

3.1 Feature Extraction

Feature extraction is the most important part of supervised classification and which palys a major role in medical image analysis. Extracting numerical features from the region of interest which represent the particular image is called feature vectors. Intensity of the mammogram



Figure 1. Various phases of the proposed method.

may vary which depend upon the degree of density of the breast. Therefore, Statistical features gives more useful information for intensity based variation in the images. Statistical moments mean, standard deviation, skewness and kurtosis are extracted from the region of interest.

Mean: It measures the average pixel intensity.

$$\mu = \frac{\sum_{ij} X_{ij}}{N} \tag{1}$$

Standard deviation: It measures the dispersion of a set of data from its mean

$$\sigma = \sqrt{\frac{\sum_{ij} \left(X_{ij} - \mu\right)^2}{N}} \tag{2}$$

Skewness: It measures the asymmetry of the pixel values around the image mean.

$$\frac{\sum_{ij} \left(X_{ij} - \mu \right)^3}{N\sigma^3} \tag{3}$$

Kurtosis: It measures, whether an image's intensity distribution is peaked or flat relative to the normal distribution.

$$\frac{\sum_{ij} \left(X_{ij} - \mu \right)^4}{\left(N - 1 \right) \sigma^4} \tag{4}$$

3.2 Classification

k-NN is a non-parametric method used for classification. An object is classified by a majority vote of its neigbors. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms.

- Calculate similarity between test image and each neighboring images.
- Select k nearest neighbors of a test image among training images.
- Assign test image to the class which contains most of the neighbors (majority).

k-NN uses the distance measures for finding similarity. k-NN with single distance measure is not sufficient to provide the better accuracy. Therefore k-NN with three different distances namely City-block, Euclidean and Cosine are used and majority among the three of distance output is considered as the final output. Figure 2. Shows the combined k-NN classifier.

4. Experimental Results

Figure 3 Shows the sample input ROI images. First four order statistical moments namely mean, standard deviation, skewness and kurtosis are computed from the ROI. 322 mammograms from the Mini-Mias database was taken up for this study. k-NN with city-block distance, k-NN with Euclidean distance and k-NN with cosine distance is used for classification. k-NN with Euclidean distance gave 100% accuracy for dense and fatty tissue but the k-NN with Cosine was able to classify glandular tissue better when compared to k-NN with Euclidean and City block distance. k-NN with City block gave a comparitively better performance for both dense (94%) and fatty (91.1%). The combined k-NN accuracy achieved is 91.72%. Table 1 Shows the performance of combined k-NN classifier. Table 2. Shows the comparison between proposed workdone with previous works.



Figure 2. Combined k-NN classifier.



Figure 3. Input ROI images.

Table 1. Con	mbined k-	NN cla	assifier	results
--------------	-----------	--------	----------	---------

Tissue density	D	G	F
Correct classification	112	87	97
Misclassification	0	17	9
Accuracy in (%)	100	83.65	91.51

Table 2.Comparison between our proposed workand previous work for Classification

Features	Classifier	Accuracy	Reference
Fractal features	SVM	85.7%	S. D. Tzikopoulos et al12. 2011
SIFT, LBP, texton histogram	SVM	93.548%	G. Liasis et al13. 2011
GLCM, Statistical, Histogram (ROI)	K-NN	82.5%	M. Mario et al16. 2012
Statistical moments (ROI)	Combined K-NN	91.72%	Proposed Method

5. Conclusion

The proposed work has done using MATLAB. The lower portion which is devoid of pectoral muscle and other background is chosen as the ROI and the input for the proposed system. Statistical moments are calculated from this ROI. The extracted feature vectors are given to the k-NN with three different distance measures separately. From these three outputs majority of the two outputs is considered as a final output. This combined k-NN classifier provides a better accuracy of 91.72% and may be effectively used in breast tissue classification.

6. Acknowledgement

The work has been done under University Grants Commission (UGC) Major Research Project. The financial support of UGC is greatly acknowledged with appreciation.

7. References

- Gaudin N. ARC Communications. The International Agency for Research on Cancer. World Health Organisation; 2013 Dec 12th; Lyon/Geneva. Pr223_E.pdf.
- 2. Bassett LW. Imaging the Breast. Holland-Frei Cancer Medicine. 5th ed. BC Decker Inc; 2000.
- 3. Rangayyan, Rangaraj M, Ayres FJ, Leo Desautels JE. A review of computer-aided diagnosis of breast cancer:

toward the detection of subtle signs. J Franklin Inst. 2007; 344(3):312–48.

- Leon, Stephanie, Brateman L, Honeyman-Buck J, Marshall J. Comparison of two commercial CAD systems for digital mammography. J Digit Imag. 2009; 22(4):421–23.
- 5. Bovis K, Singh S. Classification of mammographic breast density using combined classifier paradigm. Proceedings on Medical Image Understanding and Analysis; 2002.
- Oliver, Arnau, Freixenet J, Marti R, Pont J, Perez E, Denton ERE, Zwiggelaar R. A novel breast tissue density classification methodology. Inform Tech Biomed IEEE Trans. 2008; 12(1):55–65.
- Saidin N, Mat Sakim HA, Ngah UK, Shuaib IL. Computer aided detection of breast density and mass and visualization of other breast anatomical regions on mammograms using graph cuts. Computational and Mathematical Methods in Medicine. 2013. Available: http://dx.doi. org/10.1155/2013/205384.
- Sheshadri HS, Kandaswamy A. Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms. Comput Med Imag Graph. 2007; 31:46–8.
- Tagliafico A, Tagliafico G, Tosto S, Chiesa F, Marinoli C, Derchi LE, Calabrese M. Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one. Breast. 2009; 18(1):35–40.

- Oliver A, Llado X, Perez E, Pont J, Denton E, Freixenet J, Marti J. A statistical approach for breast density segmentation. J Digit Imag. 2010; 23(5):527–37.
- Tzikopoulos SD, Mavroforakis ME, Georgiou HV, Dimitropoulos N, Theodoridies S. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. Comput Meth Programs Biomed. 2011; I02:47–63.
- Oliver A, Freixenet J, Marti R, Pont J, Perez E, Denton ERE, Zwiggelaar R. A novel breast tissue density classification methodology. IEEE Trans Inform Tech Biomed. 2008; 12(1):55–65.
- Liasis G, Pattichis C, Petroudi S. Combination of different texture features for mammographic breast density classification. IEEE 12th International Conference on Bioinformatics and Bioengineering (BIBE); 2012. p. 732–7.
- 14. Subashini TS, Ramaingam V, Palanivel S. Automated assessment of breast tissue density in digital mammograms. Comput Vis Image Understand. 2010; 114:33–43.
- 15. Suckling J, Parker J, Dance D, et al. The mammogram image analysis society digital mammogram database. exerpta medica. Int Congr. 1994; 1069:375–8.
- Mustra, Mario, Grgic M, Delac K. Breast density classification using multiple feature selection. Automatika: casopisza automatiku, mjerenje, elektroniku, racunarstvo i komunikacije. 2012; 53(4):362–72.