QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Computer Aided Diagnosis System for Breast Cancer using ID3 and SVM Based on Slantlet Transform

Mohammed Salih Mahdi
BIT, Business information College, University of Information Technology and Communications- Iraq
Ms.mahdi99@gmail.com

ARTICLE INFO

Article History:

Received: 20 March 2017

Accepted: 1 April 2017

Published: 10 April 2017

DOI:

10.25212/lfu.qzj.2.2.16

Keywords: *Breast cancer, CAD, Slantlet, ID3, SVM*

ABSTRACT

Lately, the woman's chest cancer is the 2nd reason of deaths in females. Mammogram images are medical images, which can be read by physicians to detect breast carcinomas. In this paper, proposed Computer Aided Diagnosis system that can assist the doctors in hospitals to improve the diagnosis of the disease to detect cancer cells. Enhance the undesirable effects the of mammogram images by using slantlet transformer , set of different stages and classifies as normal, abnormal according to ID3 and SVM. For the same testing set, the practical outcomes displays SVM classifier with an accuracy of 95% and ID3 classifier with an accuracy of 92% based on MIAS database.

INTRODUCTION

The Medical images helpful in the education area for learners by clarifying these images will assist them in their projects [1]. The explanation of medical images acts the most significant and exciting part of technology. For many experts, aiding physicians is becoming of high and advantage value by evolving a computer assisted diagnosis for carcinoma [2]. Mammogram images are top-resolution x-ray imaging of breast. This includes ray transmission during tissue and the dropping of anatomic structures on a film display or image sensor. A linked with the x-ray imaging dropping is a decrease in anatomic data into a 2D image. Studies have display that the death rate could decrease by 34% if all females age fifty and older have regular mammogram [3]. In exercise of mammography imaging, there exist essentially two kind of normal tissue differentiable in the images normal and abnormal. A sample mammogram displaying the breast anatomy is shown in Figure.1 [4].

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

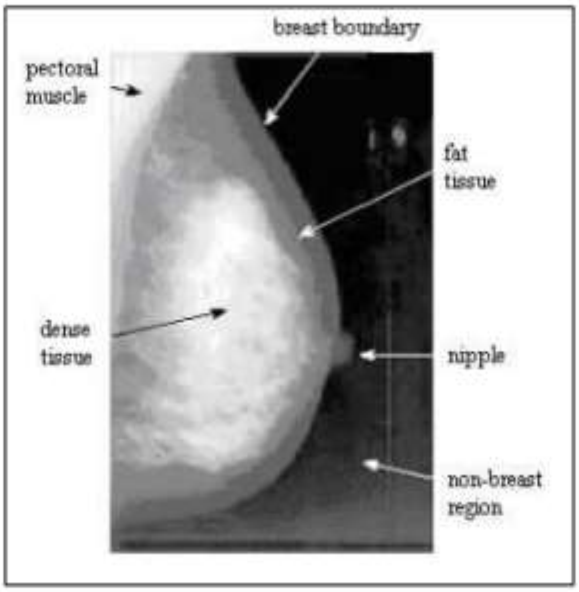


Figure 1 :Mammographic breast Anatomy[4]

Some proposals have been drawn from the scientific literature: Sara D. And Abbasi, M. [2], breast cancer is scourge diseases amongst females, needing CAD system in diagnosing tumors of breast cancer and helping medical technicians using classifier based on SVM. Subbiah B. [5], reduce the number of false positives and false negatives. by implemented a CAD that can help the radiologists. Leena J. Baskaran S. and Govardhan A. [6] building a CAD for classification in digital mammograms using contourlet transform and SVM.

THEORY

The transformation is a procedure that converts an object from scope to another scope so as to have some paramount implied data, that used for CAD system [7] The aim of transformation is to replace the data from time-space scope to time-frequency scope, that produce leading compression outcomes [8]. The information on spatial scope is generally less fixed than frequency scope. Thus, presented lazy and complex in spite of providing higher accuracy [9].

Slantlet transform (ST) is founded based on an improved version of the usual discrete wavelet transform filter bank where the support of the discrete-time basis functions is reduced. However, Slantlet transform (ST) is according to an amended version of the discrete wavelet transform (DWT) where reducing the backing of DWT-time basis functions. For variant scales, ST designing variant filters unlike iterated filter bank technique for the DWT. For each level, It utilizes distinct filters while DWT is performed in the form of an iterative filter bank. [10]. Some feature of ST are [11]:

- ST is a multi-resolution and better time localization.
- ST better compression as compared to DCT

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

- Filter bank of ST is orthogonal.

Filter bank of ST-Level , using the down-pointing arrows (↓) as shown in Figure (2) appear down-sampling by 4 and transform matrix same size of image.

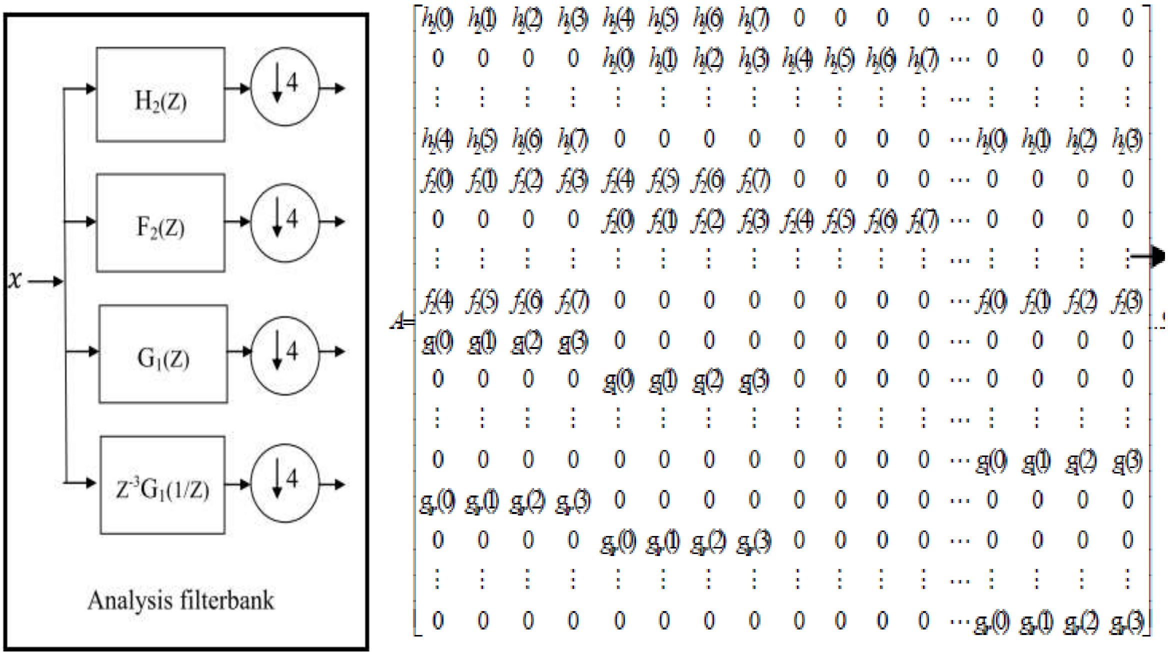


Figure 2: Analysis ST-Level2 filterbank and transform matrix.

The in equation (1,2,3,4), filters coefficients of ST level 2 filter bank can be gained as [12]:

H2= [0.27 , 0.395, 0.52, 0.645, 0.23, 0.1, -0.01, -0.148] ------------------------- (1)

F2= [-0.083, -0.12, -0.79, -0.197, 0.753, 0.344, 0.064, -0.474] ----------------- (2)

G1= [-0.512 , 0.828, -0.12, -0.195] --- --- (3)

G1r= [-0.195, - 0.12, 0.828, 0.512] -- (4)

The decision tree is a common approach for pattern classification and consider as a class-labelled. Decision tree same to tree structure, so very inner node indicates a test on an attribute, each branch acts an result of test, and any leaf node catch a class label [13] . decision tree algorithm by Quinlan Ross presented in the year of 1986. ID3 utilizes information gain measurement for choosing the attribute of splitting. It only accepts categorical attributes in constructing a model of tree. 3 widely known diversity functions are depicted below. Let there be a data-set S (training data) of C outputs. Let P (I) represent the proportion of S that belongs to a class I where I is different from one to C for the classification problem with C class [14]:.

Simple Diversity index = p (I)--(5)

The goodness of a split measures of entropy, which measures the amount of information on a feature.

$$\text{Entropy}(S) = \sum_{I=1} (-P(I)\log_2 P(I)) ------------------ (6)$$

Gain(S,A) defined as

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum\left(\left(\frac{|SV|}{|S|}\right) \times Entropy(SV)\right) ------ \quad (7)$$

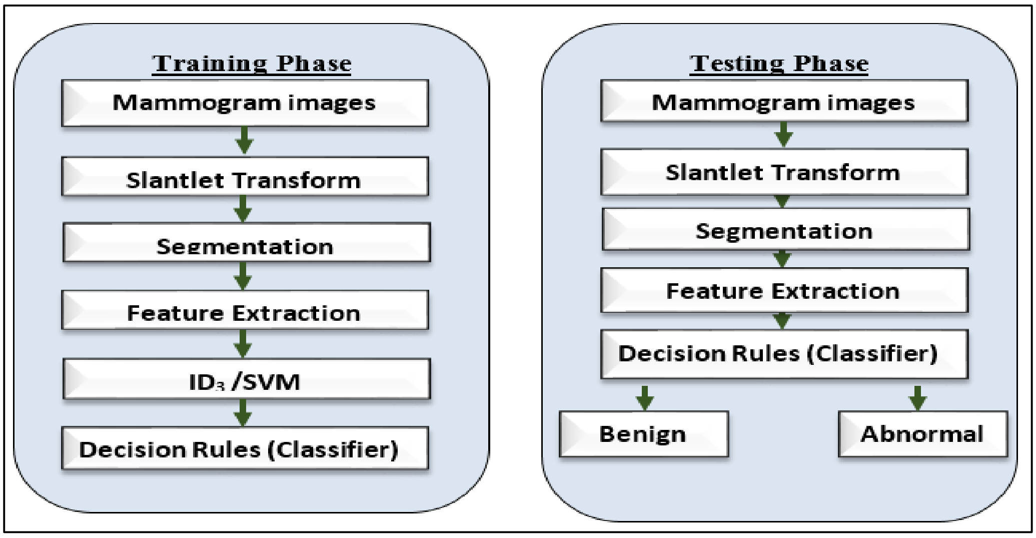
Support Vector Machines : the modern technique in machine learning is SVM in recent years, that is common in many pattern classification issus including texture classification and pattern recognition . According to different kernels SVM maximizing marginal range between classes with decision outlines drawn. Binary SVM is intended to job with only 2 class by determining hyper-plane to split the 2 class. The patterns which nearest to the margin which was chosed to define hyper-plane which known as support vectors. [15].

METHODOLOGY

The phase structure of the Computer Aided Diagnosis system for breast cancer (CAD-BC) made up of 2 phase: training and testing. Every one of the steps has predetermined functions. Every function is depicted in detail in the upcoming sub-sections which can be shown in following Figure 3. Mammography images Analysis Society (MIAS) implemented in this research was taken from , which is a UK research group organization related to the breast cancer. 250 images selected from the data base, 175 images used for training and 75 images used for testing. Enhance the undesirable effects the of mammogram images , improving the fitness of the mammogram image and produce the feature extraction step simple and extra credible by using slantlet transform with clearness , rid from noise and smooth by blocking detailed information. Region of interest(ROI) is the most important step in CAD-BC. breast object is segmented from the background using mean's thresholding segmentation. It is easy to be performed. The conception is to make a threshold which represented by the mean of all pixels of mammogram images .such that each pixel in mammogram images is more than density to threshold remained otherwise removed. extracted features are clue of ROI outcomes . For any mammogram images, Gray

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Level Co-occurrence Matrix (GLCM) is build with distance=1 and aveage direction of (0,45,90,135) degree. attributes are extracted based on GLCM are Contrast, Dissimilarity, Entropy, Homogeneity and Standard deviation. These attributes are the input of the classifier model . The classifier model built by using ID3 and SVM. The classifier model to test a new woman's chest image which is not exist in training set.



A- B-

Figure 3: The structure of CAD-BC: A- training of CAD-BC. B- testing of CAD-BC.

RESULTS

The performance of the CAD-BC is estimated by using confusion matrix, running time and classification accuracy which has been obtained from the testing part by using slantlet transforms can be shown in a table (1) , table (2) and table (3). Also, images quality after slantlet transforms measured by PSNR can be shown in a table (4). Table (5). Shows the Comparison of CAD-BC with related work system and finally illustrated the simple images of stages of mammogram processing in figure 4.

Table 1: Confusion matrix using ID3.

Actual Class	Predicate Class	
	Normal	Abnormal
Normal	43	2
Abnormal	4	26

Table 2: Confusion matrix using SVM.

Actual Class	Predicate Class	
	Normal	Abnormal
Normal	42	3
Abnormal	1	29

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Table 3: Comparison between ID3 and SVM

Class Type	ID3		SVM	
	Running time	Accuracy	Running time	Accuracy
Average	90 second	92%	116 second	95%

Table 4: Performance measures of the Slantlet transforms.

Class Type	Slantlet transforms
	PSNR
Normal	31.052
Abnormal	30.061

The implementation of CAD-BC compared with other related system with same data set (MIAS database) as following table.

TABLE 5: Comparison with related work system

Item	Accuracy
Sara D. and Abbasi	87%
Subbiah B.	91
Leena J. Baskaran S. and Govardhan A	93%
Mohammed salih (CAD-BC)	92 % ID3
Mohammed salih (CAD-BC)	96% SVM

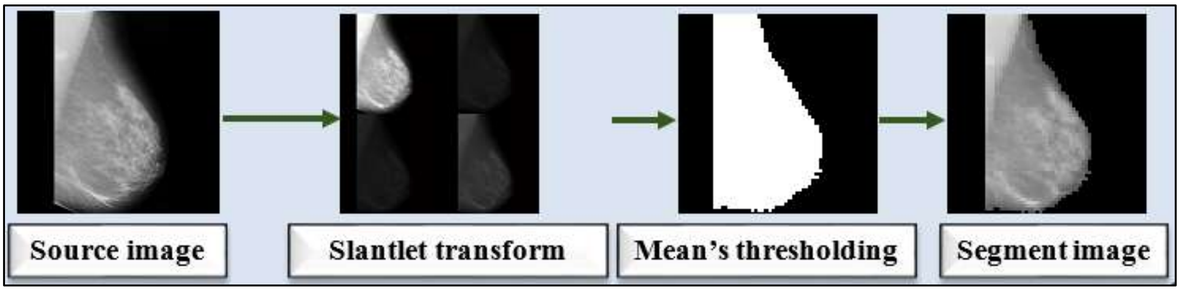


Figure 4: Stages of mammogram processing

CONCLUSION and RECOMMENDATIONS

True indicator of breast cancer diseases are mammogram images. Automated segmentation using mean's threshold, which extracted the object (breast), For the same testing set, the practical outcomes displays SVM classifier with accuracy of 95% and ID3 classifier with accuracy of 92% based on MIAS database. it is strict to assume that every individual attribute is superior to others based on outcomes . The five attributes used in this paper which can be extracted from GLCM matrix, and which contain information about image features.

Enhancing the woman's chest cancer by using different transforms like framelet. Appling a various classification techniques to build the classifier like naive , ANN.

REFERENCES

[1] Smitha P., Shaji L. and Mini M., "A Review of Medical Image Classification Techniques", International Conference on VLSI, Communication and Instrumentation (ICVCI), Karunagapally, 2011.
[2] Sara D. and Abbasi M., " Breast Cancer Diagnosis System Based on Contourlet Analysis and Support Vector Machine", Department of Computer, Science and Research Branch, Islamic Azad University, Khuzestan, Iran, 2011.
[3] Yajie S., "Normal Mammogram Analysis", PhD Thesis, Department of Computer Science, University of Purdue, 2004.
[4] Monika S., "Computer Aided Diagnosis in Digital Mammography: Classification of Mass and Normal Tissue", MSc Thesis, Department of Computer Science and Engineering, University of South Florida, 2003.
[5] Subbiah B.,"Contourlet Based Texture Analysis and Classification Of Mammogram Images", International Journal of Engineering Science & Technology, Vol. 5 Issue 6, p1228, 2013.
[6] Leena J., Baskaran S. and Govardhan A.," A Robust Approach to Classify Micro calcification In Digital Mammograms Using Contourlet Transform and Support Vector Machine", JCA (ISSN: 0974-1925), Volume VI, Issue 1,2013.
[7] Matheel E, "Color Image Denoising Using Discrete Multiwavelet Transform" , PhD Thesis, Department of Computer Science and Information Systems of the University of technology, 2000.
[8] Panrong X., "Image Compression by Wavelet Transform ", the faculty of the Department of Computer and Information Sciences, East Tennessee State University, Masters Abstracts International, Volume: 40-01, page 0197, 2001
[9] Jing Y., Yong H. and Phooi Y. ,"Recent Trends in Texture Classification: A Review", Symposium on Progress in Information & Communication Technology, pp. 63-68, 2009.
[10] Ivan W. Selesnick, "The Slantlet Transform", IEEE Transactions on Signal Processing, Vol. 47, No. 5, May, 1999..
[11] Iman M. Ga'fer and Afrah L. Mohammed, "Slantlet Transform based Video Denoising", Baghdad Science Journal, Vol. 8, No.2, 2011.
[12] N. B Patil, V. M. Viswanatha and S. Pande, "Slant Transformation as a Tool for Pre-Processing in Image Processing", International Journal of Scientific & Engineering Research, Vol. 2, No. 4, April, 2011
[13] Matthew N. and Sajjan G.," Comparative Analysis of Serial Decision Tree Classification Algorithms ", IJCSS, 2013.
[14] Lavanya V. and Rani U.,"Performance Evaluation of Decision Tree Classifiers on Medical Datasets", IJCA, 2011.
[15] C. Chen, C. Chen and C. Chen, "A Comparison of Texture Features Based on SVM and SOM," ICPR, vol. 2, pp. 630-633, 2006.