International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 9, Volume 2 (September 2015) www.ijirae.com

Classification of Mammography Image Using Machine Learning Classifiers and Texture Features

Hashem B. Jehlol*
Al-Mustansiriyah University
Computer Center

Zainab khyioon Abdalrdha
Al-Mustansiriyah University
College of Education
Department of Computer Science

Anwer Subhi Abdulhussein Oleiwi
Al-Mustansiriyah University
Avicenna Center for E-learning

Abstract— Research in the field of classification of mammography is very important because it is best methods used to discover breast cancer at an early stage. The purpose of currant research is to propose ways to automate the process of mammography classification. This requires a description of the image using the feature extraction algorithms and then classified using machine learning algorithms. First and Second Order Statistics texture descriptors were used to describe mammography images. Several machine learning algorithms were used to classify images such as Random forest (RF), The Naive Bayes (NB), C4.5, The multi-layer perceptron (MLP) and Decision Table (DT). The goal is to find the best combination for feature extraction algorithm and classification algorithm, which gives good results in the classification of mammograms. The best results were obtained when the mammography images described using Second-Order Statistics together with Random forest as a classification technique.
Keywords— mammography, Breast cancer, Random Forest, Naive Bayes, C4.5

I. INTRODUCTION

One of the most prevalent types of cancer among women is breast cancer, according to statistics of the World Health Organization reports. Approximately one in eight women has a risk of developing breast cancer in the United States [1].Chances of survival increase if the cancer is detected at an early stage. There are three methods used to detect breast cancer: mammography, biopsy and biopsy needle. The first step towards the detection of breast cancer is mammography [2].A mammogram is an X-ray system to check the breast. X-ray mammography is currently considered as standard procedure for diagnosis breast cancer. The diagnosis result of tissue is classified into three categories: Normal, benign and malignant. Normal represents mammogram without any cancerous cell, benign represents mammogram showing a tumour, but not produced by cancerous cells and malignant represents mammogram showing a tumour with cancerous cells. It is difficult to distinguish a benign micro-calcification from malignant [3].

In an attempt to improve early detection of breast cancer, our research focuses on the conduct mammography X-ray analysis, which aim to improve the differentiation of benign from malignant cases, or the detection of suspicious cases in general. Therefore we compare several techniques for automated mammography image classification. The organization of this paper is as follows. A brief overview about previous research in the field of detection and classification of breast cancer is presented in Section 2. Section 3 presented the proposed methods that used in the research. Section 4 presented an overview of the experimental setup. Section 5 contains experimental results and discusses the results. Finally conclusions are presented in section 6.

II. RELATED WORK

Many computer algorithms conducted to detect and classify suspicious regions in digital mammogram. These algorithms are estimating the probability of malignancy based on features derived from regions of interest (ROI) for a given mammogram. Most of them direct towards classifying a mammogram into normal, benign or malign. Some of these algorithms are discussed below.

J.S. Leena Jasmine et al. Proposed method aims to use features such Gabor features, curvilinear features, texture features and multi resolution features for the purpose of detection and classification of breast cancer using mammographic images. They used support vector machine. In their study they obtained accuracy was 91.4%. The DDSM database mammograms images were used [4].

De Melo et al. proposed method for determining the set of features that are used to build a best automatic classification. The scalar feature selection is used to generate sets which are consisted of different numbers of features. The distance measurements such as the area under receiver operating curve and fisher's discriminant ratio were used. For classification purposes, different types of architectures of feed forward neural networks were used [5].

Aswini kumar mohanty et al. proposed method aims to apply image mining for breast mammograms to detect and classify the cancerous tissue. They present method of feature selection that reduces about 60% of association rule and the features. A total of twenty six features including GLCM features and histogram intensity features were extracted. A dataset of images consisting of 322 images taken from a MIAS dataset was used in the experiment. The accuracy obtained by applying this method is about 97.7% [6].

S. Shanthi et al. proposed method to evaluate the texture classification of mammographic when using features extracted from ridgelet transforms, wavelet and co-occurrence matrices. They used database consisted of 120 mammographic images, half of them abnormal images and other half normal images. The texture descriptors (Entropy, sum variance, sum average, cluster tendency and energy) were calculated to analyse ROI texture patterns [3]

Herwanto et al. proposed method to classify mass and micro-calcification in mammogram using association technique. This method consists of three phases (1) a pre-processing phase to enhance image quality and followed by segmenting ROI, (2) a phase for mining a transactional table and (3) a phase for organizing the resulted association rules in a classification model. Mammographic MIAS dataset was used to evaluate the proposed method. The mean features and GLCM proved to be possibility for distinguishing micro-calcification from mass. When applying this method it was obtained accuracy of 83% [7]

S shanthi et al. proposed a feature set based on fractal analysis, Gabor filters and multi scale surrounding region dependence method (MSRDM) to identify the most common appearance of breast cancer namely micro-calcification, masses and architectural distortion. The results of the experiments indicate that the proposed features with SRAN classifier can improve the classification performance. The SRAN classifier produces the classification accuracy of 98.44% for the proposed features with 192 images from MIAS dataset [8].

III. Proposed Method

To carry out experiments, two types of feature extraction algorithms and five methods of classification were used to compare and find the best suitable combination. Fig.1 illustrates the methodology of the mammogram image classification:

Digitizing mammogram
↓
Region of Interest (ROI)
↓
Image Prepressing
↓
Feature Extraction
↓
Classification
↓
Evaluation

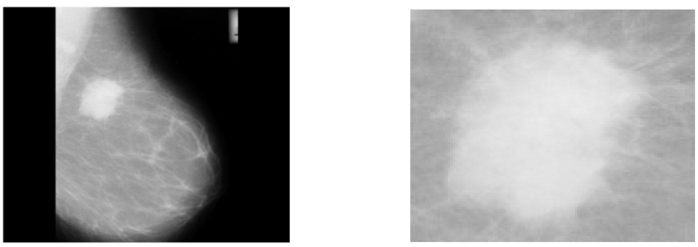
Fig. 1: Framework of proposed approach

A. *Digitizing Mammogram*

Acquisition of digital mammogram images represents the first step in the proposed method. Mammography Image Analysis Society (MIAS) database [9] has been used to conduct the experiment in this paper. It has 322 images, which belong to three categories, namely: (1) Normal is a collection of mammogram images that do not have breast cancer, (2) Benign, is abnormal mammogram images have a benign breast cancer, and (3) malignant, is abnormal mammogram images have Invasive breast cancer. The resolution of each mammogram images was 1024x1024 while the accuracy of grey level was eight-bit [10]. The proposed approach focuses on images classification to abnormal (benign, malignant) and normal.

B. *Reign of Interest*

After collecting mammogram images, the next step is to determine the region of interest ROI for mammogram images. ROI extracted by entering coordinates X, Y and radius in pixels, according to data provided by the MIAS database for each abnormal mammogram image. A random 60x60 pixels region was extracted for the normal mammogram images. Fig. 1 illustrates the extracted portion of ROI mammogram images with malignant mass.

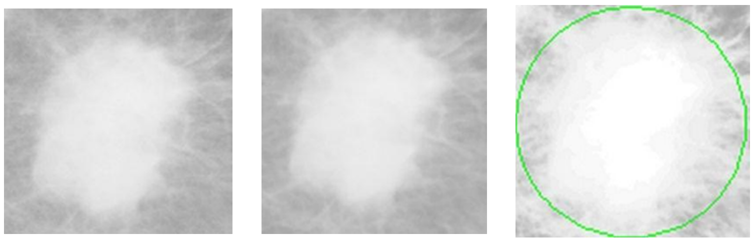


(a) (b)

Fig. 1: a) original mammogram image , b) The Region of Interest

C. Pre-processing of Mammogram Images

The main problem facing extracting the visual features of the mammographic images is noise, the different resolution, quality and the weak contrast of the mammograms. This makes the detection of the cancer much harder. Pre-processing is required to overcome this problem and make efficient feature extraction of images as possible. Median filter was used to remove the noise from the mammogram images. When using median the value of an output pixel is determined by the median of the neighbourhood pixels. Median filtering has the ability to remove noise without reducing the image sharpness, as shown in fig. 1a and fig. 1b. The histogram equalization was used to adjust the contrast of the mammogram images. This method improves the contrast by spreading out the most frequent intensity values. Histogram equalization was used to make contrast adjustment so the anomalies can be better emphasized [11], as shown on fig. 1c.



(a) (b) (c)

Fig.2: a) The Region of Interest, b) ROI after used median filter c) ROI after used histogram equalization

D. Feature Extraction Methods

In this paper two common feature extraction algorithms were used theses algorithms based on first and second order Statistics to compare their performance.

1) *First-Order Statistics*: First-order Statistics features are computed from the original values of image. In first-order features the relationships with neighbour pixels do not consider. This approach gives Many features include moments of an image Ii(x,y) with the size M×N. These features include mean, standard deviation, kurtosis, skewness, and entropy such as in the following equations [12].

$$mean(u) = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} Ii(x,y)}{M*N} \quad (1)$$

$$standard\ deviation\ (\sigma) = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N} Ii(x,y) - u}{M*N}} \quad (2)$$

$$kurtosis = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} Ii(x,y) - u^4}{M*N*\sigma^4} \quad (3)$$

$$skewness = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} Ii(x,y) - u^3}{M*N*\sigma^2} \quad (4)$$

$$entropy = \frac{1}{M*N}\sum_{x=1}^{M}\sum_{y=1}^{N} Ii(x,y)(-\ln Ii(x,y)) \quad (5)$$

2) *Second-Order Statistics*: The second-order statistics based on the relation between two neighboring pixels in one offset, where the first pixel is called the reference and the second the neighbor pixel. The Grey-Level Co-occurrence Matrix (GLCM)) is used to estimate image properties related to second-order statistics [13].A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the image. The matrix element P (i, j | Δx, Δy) is the relative frequency with which two pixels, separated by a pixel distance (Δx, Δy), occur within a given neighborhood, one with intensity 'i' and the other with intensity 'j'. The matrix element P (i, j | d, ɵ) contains the second order statistical probability values for changes between gray levels 'i' and 'j' at a particular displacement distance d and at a particular angle (ɵ). A displacement, d could take a value of 1, 2, 3… n whereas an angle, ɵ is limited 0°, 45°, 90° and 135° [14].

In our experiments the GLDM feature descriptor was calculated from five values of the angle ɵ (0°, 45°, 90° and 135°), five texture features (Energy, Contrast Correlation, Homogeneity, Entropy) and displacement distance d=1 thus, the implementation has 20 features. The texture features are calculated as follow:

Energy: Energy represents regularity in the mammographic image. Energy in general is calculated from the value of the mean squared signal. It calculated according to the following equation **[**15**]**.

$$energy = \sum_{i,j=0}^{n-1} P(i,j)^2 \qquad (6)$$

Contrast: the contrast measures the variation between the lowest and the highest values of a neighboring set of pixels. It measures the amount of existing local differences in the image [15].

$$contrast = \sum_{i,j=0}^{n-1} (i-j)^2 \, P(i,j) \qquad (7)$$

Correlation: The correlation returns a measure of how a pixel is correlated to its neighbor over the whole image **[**16**]**.

$$correlation = \sum_{i,j=0}^{n-1} \frac{(i \times j)\, P(i,j) - u_i u_j}{\sigma_i \sigma_j} \qquad (8)$$

σ2 = the variation of the intensities of all reference pixel in the relationships that contributed to the GLCM, calculated as:

$$\sigma^2 = \sum_{i,j=0}^{N-1} P_{ij}\,(i-u)^2 \qquad (9)$$

Homogeneity, Angular Second Moment (ASM): ASM used to measure the homogeneity of the image [15].

$$homogeneity = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \{P(i,j)\}^2 \qquad (10)$$

Entropy**:** Entropy refers to the measure of the disorder or complexity in the image. The highest value of Entropy is found when the values of P (i, j) are allocated quite uniformly throughout the matrix. Entropy is strongly but inversely correlated to Energy [15].

$$entropy = -\sum_{i,j=0}^{n-1} P(i,j) \log P(i,j) \qquad (11)$$

Where '*i*' indicates the rows of the GLCM matrix, 'j' indicates the columns of the GLCM matrix, 'n' indicates the number of gray levels and *P(i, j)* is the cell denoted by the row and the column of the GLCM matrix [17].

E. Classification of mammograms

There are various algorithms for automated classification. In this paper used several classification algorithms to compare their performance: Random forest (RF), The Naive Bayes (NB), C4.5, The multi-layer perceptron (MLP) and Decision Table (DT).

Random Forest (RF) is an approach which has been proposed by Breiman for classification tasks. It mainly comes from the combination of tree-structured classifiers with the randomness and robustness provided by bagging and random feature selection [18]. The classification is performed by sending a sample down is each tree and assigning it the label of the terminal node it ends up in. At the end the average vote of all trees is reported as the result of the classification. Random forest is very efficient with large datasets and high dimensional data [2].

The Naive Bayesian (NB) is based on the Bayesian theorem [19].The Naïve Bayesian Classifier assumes that features are independent [20].This method is important for several reasons. It is very easy to construct, dose not need any complicated iterative parameter estimation schemes. This means it may be readily applied to huge data sets [21].This classification technique analyses the relationship between each attribute and the class for each instance to derive a conditional probability for the relationships between the attribute values and the class [19].

C4.5 is an extension of ID3 algorithm that was designed by Quinlan to deal with issues that cannot be handled by the ID3 algorithm. These include avoidance of over fitting the data; reduced error pruning, rule post-pruning, handling continuous attributes and handling data with missing attribute values [1]. It attempts to build a decision tree with a measure of the information gain ratio of each feature and branching on the attribute which returns the maximum information gain ratio [22]. Pruning takes place in C4.5 by replacing the internal node with a leaf node thereby reducing the error rate [23].

The multi-Layer Perceptron (MLP) is a feed forward neural network consisting of an input layer of nodes, followed by two or more layers of perceptron, the last of which is the output layer. The layers between the input layer and output layer are referred to as hidden layers [23]. The major aim of MLP algorithms is to automatically learn and make intelligent decisions. It is known as feed forward because it does not contain any cycles and network output depends only on the current input instance. Learning take place by changing connection weights after each piece of data is processed, based on the amount of error in the target output as compared to the expected result. [24].

A Decision Table (DT) is the method used to build a complete set of test cases without using the internal structure of the program in question. In order to create test cases we use a table to contain the input and output values of a program. Such a table is split up into four sections [25].Two variants of decision table classifiers are available. The first classifier, called DTMaj (Decision Table Majority) returns the majority of the training set if the decision table cell matching the new instance is empty, that is, it does not contain any training instances. The second classifier, called DTLoc (Decision Table Local), is a new variant that searches for a decision table entry with fewer matching attributes (larger cells) if the matching cell is empty. This variant therefore returns an answer from the native region [26].

IV. EXPERIMENTAL

To conduct experiments in the proposed method, MIAS database was used. The MIAS database was created to contain two experimental datasets on the same images. The difference between them is that in the first dataset the images are split in two classes: normal or abnormal, and in the second dataset the images are split in three classes: benign, malign and normal. MIAS database is a set of 322 commented images. The abnormal images in this database contain the coordinates and the radius. Matlab 2010 was used to extract all features methods, the process of median filter and histogram equalization process that implemented in this research. The Weka library was used for all classification methods that implemented in this method. The following steps briefly describe the experiment in this research:

Setep1: from all mammogram images in the dataset the region of interest was extracted. The region of interest was extracted from the abnormal images depending on the information contained in the dataset while a random region of 60x60 pixels was extracted for the normal mammogram images.

Setep2: To remove noise from mammogram images and improve the quality of the region of interest the median filter and histogram equalization were applied.

Setep3: the features were extracted from the normalized image regions using First-Order Statistics and second-Order Statistics.

Setep4: Mammogram images were classified using number of classification algorithm such as (Random forest, The Naive Bayes (NB), C4.5, The multi-layer perceptron (MLP) and Decision Table (DT) classifier).

Setep5: Weka tools were used for images classification with 50% percentage split. In 50% percentage split, 50% of the samples are used in the training phase and the remaining samples are used in the testing phase.

The procedure was performed in the two-case (normal or abnormal) and the three-case (benign, malign and normal) dataset using all previous steps. The goal of the study is to conclude which combination of classification method and feature extraction method will yield the best results.

V. RESULTS AND DISCUSSION

Table 1 show the accuracy and model build time of the classification for the two-case (normal and abnormal). The table is divided in two parts. The first half describes the results when the features are extracted only from first order features descriptor and the second half describes the results when the features are extracted from second-order Statistics features. The best result of classification accuracy was 98.8% and the model build time was about 0.14 second in the case of the random forest classifier and when the mammogram images are described using second-order statistics (GLDM). The best results of NB, C4.5, DT and MLP classifiers are also in the case of the second-order statistics descriptor with 97.8%, 97%, 95% and 94% classification accuracy, respectively such as shown in Fig.3.

Table 1: Classification accuracy for two-case (normal and abnormal)

Descriptor/ Classifier	First-order Statistics		second-order Statistics	
	accuracy	Model Build Time(second)	accuracy	Model Build Time(second)
RF	96.8	0.13	98.8	0.14
NB	94.5	0.02	97.8	0.2
C4.5	95.6	0.05	97	0.05
MLP	93.9	1.27	94	7.88
DT	94.5	0.02	95.9	0.03

When using first and second order statistics features the MLP takes more time to classify mammogram images for normal and abnormal than other classifiers. Can note the difference in the time required for the classification of mammogram images due to the interior design of each classifier. Fig.4 show the classifiers used in this paper and the time required for each one.

International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 9, Volume 2 (September 2015) www.ijirae.com

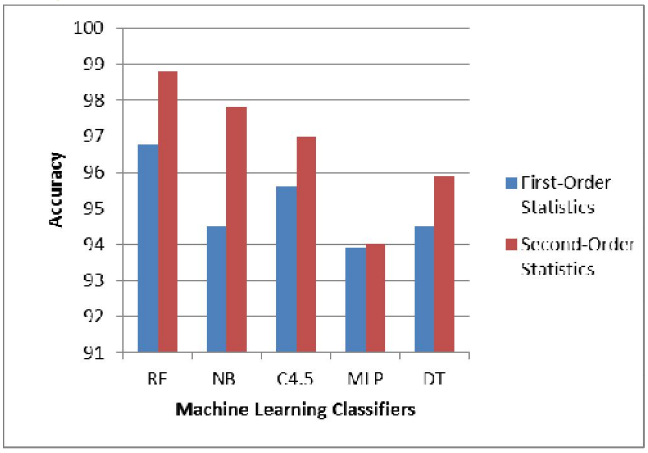


Fig.3 the accuracy of the classification for two-case (normal and abnormal) when using first and second order statistics features

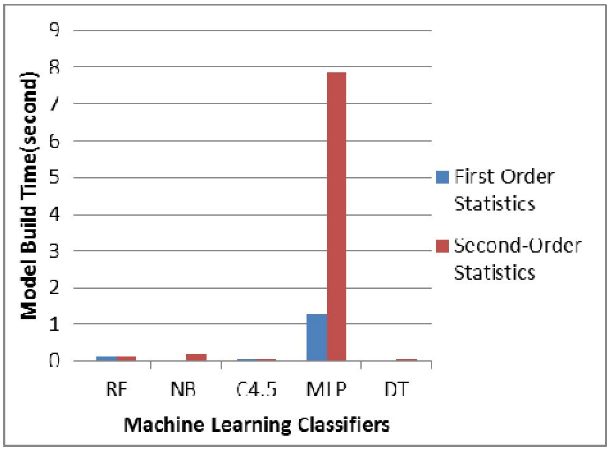


Fig.4 the time for each classifier to classify two-case (normal and abnormal) when using first and second order statistics features

The results for the three-case (benign, malign and normal) are presented in Table 2. In the state of the three-case the results are similar to the two-case and the best result of classification performance was in the state of the random forest classifier when the images are described using second-order statistics descriptor with a classification accuracy of 85.8% and model build time 0.14 second as shown in Fig.6.

Table 2: Classification accuracy for three-class (benign, malign and normal)

Descriptor/ Classifier	First-order Statistics		second-order Statistics	
	accuracy	Model Build Time(second)	accuracy	Model Build Time(second)
RF	83	0.12	85.8	0.14
NB	82	0.04	83.4	0.06
C4.5	83	0.03	84	0.03
MLP	82	1.22	83	8.09
DT	83.3	0.05	84.2	0.04

In three-case again, the MLP classifier takes more time to classify mammogram images than other classifiers. Fig.6 show the classifiers used in this paper and the time required for each one.

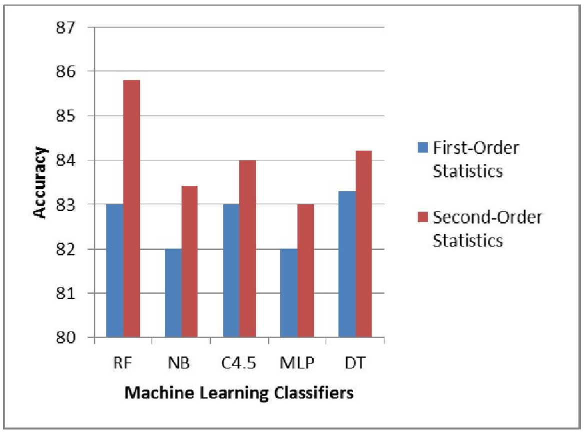


Fig.6 the accuracy of the classification for three-case (benign, malign and normal) when using first and second order statistics features

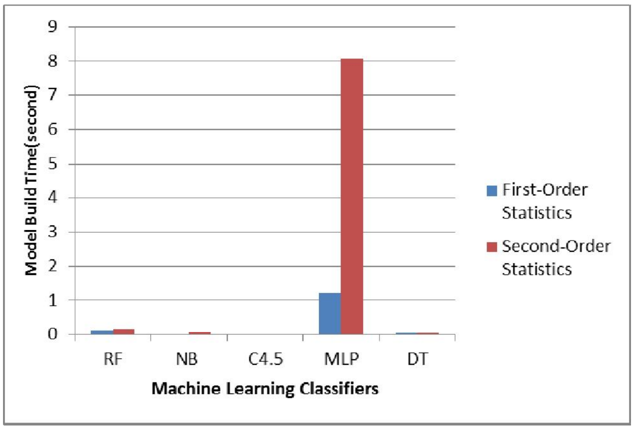


Fig.6 the time for each classifier to classify three-case (benign, malign and normal) when using first and second order statistics features

The Experience demonstrated superiority second order statistics features GLDM descriptor due to different angle has been combined in one feature vector. For every angle, ɵ (0°, 45°, 90° and 135°) the GLDM used five features (Energy, Contrast Correlation, Homogeneity and Entropy) and then sequencing them in one feature vector and thus implementation has 20 features. This means that the image is described using GLDM give richer description of image because it describes the different direction of textures appearing in the images. In case of first-order statistic features the relationships with neighbour pixels do not consider therefor it provides less description of the image then second-order statistic features (GLDM).

VI. CONCLUSION

Breast cancer is one of the most common cancers among women around the world. In this paper, first-order statistics and second-order statistics texture descriptors were used to describe mammographic images. Then, the resulting descriptors were classified using five classification algorithms. The best results were achieved in the case of using combination of random forest classifier and the second-order features with 98.8% classification accuracy. It can be noted that the first-order features descriptor provides poorer image description. The classification with high accuracy can support the radiologists to take an accurate diagnostic decision without more unnecessary biopsies.

REFERENCES

[1] K.Rajesh, S. Anand," Analysis of SEER Dataset for Breast Cancer Diagnosis Using C4.5 Classification Algorithm ", International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 2, 2012.
[2] Blagojce Jankulovski, Ivan Kitanovski, Katarina Trojacanec, Ivica Dimitrovski, Suzana Loskovska, "Mammography Image Classification Using Texture Features " , The 9th Conference for Informatics and Information Technology , 2012.
[3] J.S. Leena Jasmine, S. Baskaran and A. Govardhan, "A Robust Approach to Classify Microcalcification In Digital Mammograms Using Contourlet Transform And Support Vector Machine ", American Journal of Engineering and Applied Sciences, 6 (1): 57-68, 2013.
[4] Wang, D., L. Shi and P.A. Heng, "Automatic Detection of Breast Cancers in Mammograms Using Structured Support Vector Machines". Neurocomputing, 72: 3296-3302. DOI: 10.1016/j.neucom.2009.
[5] De Melo, C.L.S., C.F.F. , CostaFilho, M.G.F. Costa and W.C.A. Pereira, " Matching input variables sets and feed forward neural network architectures in automatic classification of microcalcifications and microcalcification clusters" , Proceedings of the 3rd International Conference on Biomedical Engineering and Informatics, Oct. 16-18, IEEE Xplore Press, Yantai, pp: 358-362. DOI: 10.1109/BMEI, 2010.
[6] Aswini kumar mohanty, Sukanta kumar swain ,Pratap kumar champati ,Saroj kumar lenka , "Image Mining for Mammogram Classification by Association Rule Using Statistical and GLCM features" , IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, 2011.
[7] Herwanto, Aniati Murni Arymurthy, "Association Technique Based On Classification For Classifying Micro-calcification And Mass In Mammogram "IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013.
[8] S. Shanthi, V. Murali , "A Novel Approach for Classification Of Abnormalities In Digitized Mammograms " ,Indian Academy of Sciences , Vol, 39, Part 5, October 2014.
[9] Karmilasari, Suryarini Widodo, Matrissya Hermita, Nur Putri Agustiyani, Yuhilza Hanum, " Sample K-Means Clustering Method for Determining the Stage of Breast Cancer Malignancy Based on Cancer Size on Mammogram Image Basis ", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 3, 2014.
[10] S.Krishnaveni1, R.Bhanumathi2, T.Pugazharasan3, "Study of Mammogram Micro-calcification to Aid Tumor Detection Using Naive Bayes Classifier ", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 3, March 2014.
[11] Test.
[12] N. Puviarasan, R. Bhavani, "Retrieval of Images Using Cascaded Features ", Journal of Engineering Research and Applications, Vol. 4, Issue 3, 2014.
[13] Miroslav Benco, Robert Hudec, Patrik Kamencay, Martina Zachariasova and Slavomir Matuska1, "An Advanced Approach to Extraction Of Colour Texture Features Based On GLCM ", International Journal of Advanced Robotic Systems, 2014.
[14] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar, "Image Texture Feature Extraction Using GLCM Approach ", International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013.
[15] F. R. Renzetti, L. Zortea, "Use of a gray level co-occurrence matrix to characterize duplex stainless steel phases microstructure ", 2011.
[16] Redouan Korchiynel, , Sidi Mohamed Farssi , Abderrahmane Sbihi , Rajaa Touahni, Mustapha Tahiri Alaoui, "A Combined Method of Fractal And Glcm Features For Mri And Ct Scan Images Classification ",Signal & Image Processing : An International Journal (SIPIJ) Vol.5, No.4, August 2014 .
[17] S.Krishnaveni1, R.Bhanumathi2, T.Pugazharasan3, "Study of Mammogram Micro-calcification to aid tumor detection using Naive Bayes Classifier ", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,2014.
[18] Mariana R. Mendoza, Guilherme C. da Fonseca, Guilherme Loss-Morais, Ronnie Alves, Rogerio Margis, Ana L. C. Bazzan , " Predicting Human MicroRNA Target Genes with a Random Forest Classifier ", plos , 2013.
[19] R.K.Kavitha1, Dr. D.DoraiRangasamy2, "Predicting Breast Cancer Survivability Using Naïve Baysein Classifier and C4.5 Algorithm ", Elysium journal Engineering Research and management, 2014.
[20] Lavneet Singh, Girija Chetty, "A Comparative Study of MRI Data using Various Machine Learning and Pattern Recognition Algorithms to Detect Brain Abnormalities " , Proceedings of the Tenth Australasian Data Mining Conference 2012.

[21] XindongWu , Vipin Kumar , J. Ross Quinlan , Joydeep Ghosh, "Top 10 algorithms in data mining" , Springer-Verlag London Limited,2007.
[22] Lipo Wang , Xiuju Fu, "Data Mining with Computational Intelligence", Springer-Verlag Berlin Heidelberg, 2005.23
[23] G. Sujatha , Dr. K. Usha Rani , "Evaluation of Decision Tree Classifiers on Tumor Datasets", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2013.
[24] Komal Sharma, AK winder Kaur, Shruti Gujral, "Brain Tumor Detection based on Machine Learning Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 103 – No.1, October 2014.
[25] Vijayarani, M. Muthulakshmi, "Evaluating The Efficiency Of Rule Techniques For File Classification",International Journal of Research in Engineering and Technology, eISSN: 2319-1163 | pISSN: 2321-7308, 2013.
[26] Vasileios A. Tatsis, "Evaluating data mining algorithms using molecular dynamics trajectories", Int. J. Data Mining and Bioinformatics, Vol. 8, No.2, 2013.