Mammogram Image Retrieval using IPSO Optimized Anfis Classifier

Sonia Jenifer Rayen, R.Subhashini

ABSTRACT-: Content-based image retrieval (CBIR) is an research area over the past years that has attracted research. In various medical applications like mammogram analysis CBIR techniques helps the medical team to get similar set of images from a large medical records to help in diagnosis of a disease. This paper proposes an efficient Content-Based Mammogram Image Retrieval method by using an Optimized Classifier. Initially, the input dataset is preprocessed, in which noise removal and contrast enhancement are done. Next, pectoral muscles of the mammogram images are removed using Single Sided Edge Marking (SSEM). Now, feature extraction is done, in which GLCM features, Gabor features and the Local Pattern with Binary features are being removed. The features that are being removed are classified into three classes namely benign, malignant and normal. An optimized classifier named as Adaptive Neuro Fuzzy Inference System (ANFIS), which is optimized by using the Improved Particle Swarm Optimization (IPSO) technique, is used for classification purpose. Finally, similarity is assessed between the trained feature distance vectors and the feature distance vectors of the input query image. Similarity assessment is done using Euclidean Distance metric and the image that has the lowest distance compared with the query is retrieved. The experimental results are obtained for the proposed system and they are compared with the existing techniques.

Keywords: Mammogram, Particle Swarm Optimization, Euclidean Distance, Gabor features, Image Retrieval and Adaptive Neuro Fuzzy Inference System.

I. INTRODUCTION

Cancer is a problem of significance as it is causing social and financial implications to health of the public. However, breast cancer is an major health related issue among the women society [1]. In a country like India, patients suffering from breast cancer will double by 2025 [2]. Early stage of identification and detection of breast cancer is important as it can substantially reduces the mortality rates [3]. Although screening mammography is an cost-effective tool that is used to detect breast cancer at early stages, nationwide screening programs result in high false-positive and false-negative rates [4].

When a new case history is needed, radiologists can prefer to use the CAD (computer-aided diagnosis) tools for identification and detecting the abnormalities [5]. Content based image retrieval systems (CBIR) are mostly being included in CADx to identify and retrieve the most similar cases to the query . CBIR has two steps namely indexing and retrieval [6].However, Radiologists often find it hard to determine whether mammogram image shows a normal or an malignant state. CBIR systems can efficiently aid the radiologists in diagnosing the breast cancer by providing a set of similar pathologies from the database [7]. CBIR is a method that can be used for retrieving similar images based on the visual information of query image that might help radiologist to interpret and analyze the visual content of the current medical images [8].

With the advancement in the field of medical imaging monitoring and accessing of medical images from database have become more tedious. Therefore, there is a need of expert system so as to retrieve the best and suitable results [9]. The increased quantities of images can be efficiently used in the diagnosis and treatment of patients [10]. One of the major difficulty faced by the CBIR system is that the retrieved images has to be similar to the lesion under consideration; otherwise, the radiologists may ignore them in diagnosis, thereby loosing their value [11].

For evidence-based medicine, there is a urgent need to access relevant images from a large medical database to support the diagnosis of certain diseases [12,22,23]. Mammography is a screening tool that is used to localize and identify the suspicious tissues [13]. Currently, the imaging techniques for breast examination, mammography is an effective and the only widely accepted method, is recognized as a vital standard for breast cancer detection by American Cancer Society (ACS) [14,24]. Analysis of mammogram is usually based on the radiologist subjective judgment; this process is sometimes difficult as well as inaccurate, resulting in missing lesion detection during the routine check. So, there is a need for a system for effective prediction and retrieval of mammogram images [15].

II. LITERATURE REVIEW

Vibhav Prakash Singh and Rajeev Srivastava [16] proposed an automated system and effective content based-mammogram retrieval system. Firstly, pre processing was done. Next for segmentation of the image, selective thresholds based seeded region growing algorithm was being introduced. 2-level discrete wavelet transform (DWT) was applied on the segmented region and wavelet based centre symmetric- local binary pattern (WCS-LBP) features were being taken.

Mugahed A. Al-antari et. al [17] presented a DBN-based CAD system ,so as to classify

Revised Manuscript Received on July 18, 2019.
Sonia Jenifer Rayen, Research Scholar, School of Computing Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India(e-mail: soniya@jeppiaarinstitute.org)
Dr.R.Subhashini, Professor , School of Computing , Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India.
(e-mail: subhaagopi@gmail.com)

three stages of breast tissues.

Syed Jamal Safdar Gardezi et. al [18] presented a new method for the classification of breast cancer using time series analysis.

Birmohan Singh and Manpreet Kaur [19] developed a methodology for differentiating the micro calcification clusters .The performance of the shape-based and texture-based features had been analyzed.

Ardalan Ghasemzadeh et. al [20] presented an method for classifying and detecting the masses in mammograms. They performed tenfold cross validation and analyzed the data complexity on each fold.

Chisako Muramatsu [21] presented the quantification and incorporation of subjective similarity for retrieval of visually similar images. The feasibility of determining subjective similarities with various abnormalities was discussed.

III. CONTENT BASED MAMMOGRAM IMAGE RETRIEVAL USING OPTIMIZED CLASSIFIER

Content-Based Image Retrieval (CBIR) is growing to be one of the important research areas in computer vision. This paper proposes an efficient Content-Based Mammogram Image Retrieval technique by means of an Optimized Classifier. Initially, the input dataset is preprocessed, in which noise removal and contrast enhancement are done. Improved Median Filter is used for the purpose of noise removal. Contrast enhancement is done using Contrast Limited Adaptive Histogram Equalization. Next, pectoral muscles of the mammogram images are removed using Single Sided Edge Marking (SSEM). Now, feature extraction is done, in which GLCM features, Gabor features and Local Binary Pattern features are extracted. The features that are being extracted are then classified into three classes namely benign, malignant and normal. An optimized classifier named as Adaptive Neuro Fuzzy Inference System (ANFIS), which is optimized using the Improved Particle Swarm Optimization (IPSO) technique, is used for classification purpose. Finally, similarity is assessed between the feature vectors of trained image and query image. The framework of the proposed is illustrated in Figure 1.
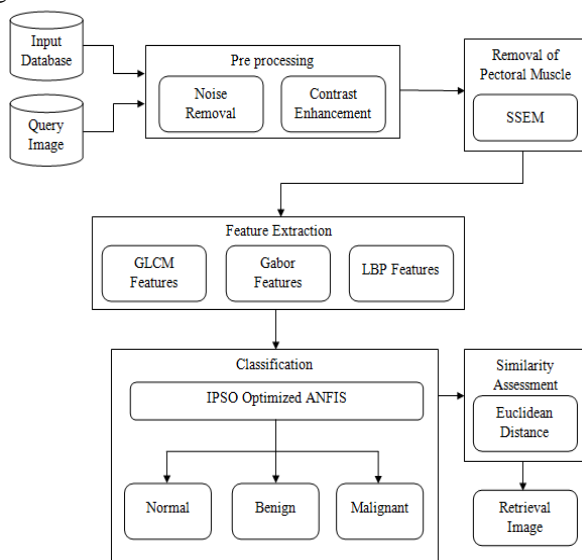

Figure 1: The Proposed Framework

3.1 Pre Processing

The preprocessing helps to enhance the quality of image, so as to limit the abnormalities and unwanted regions without having influence on background .

3.1.1 Noise Removal using Improved Median Filter

The proposed technique uses Improved Median Filter for the purpose of noise removal. This filter preserves edges while removing the noise so it is very effective . The working methodology of an median filter is that it replace each value of the image by moving through the image pixel by pixel with respect to the median value of neighboring pixels. But, in the proposed technique, standard median filter is improved and is said as Rank Order Mean. This modification consists of the exclusion of the central pixel in the window when we calculate the median. In a 3x3 window, the fifth value is taken in the ordered list as the central pixel. The output of this modified median filter is given in equation (1).

$$M(p) = \frac{r_i(p) + r_j(p)}{2} \tag{1}$$

Here, $r_i(p)$ and $r_j(p)$ are the values of the two central pixels in the ordered list of pixels.

3.1.2 Contrast Enhancement

Contrast is created by the difference in luminance reflected from two adjacent surfaces.

CL-AHE:

Contrast Limited Adaptive Histogram Equalization (CL-AHE) method is used to improve the often degraded contrast in some of mammogram images. Every block of the histogram is modified busing an transformation function given by the equation (2).

$$T_f = \sum_{i=0}^{f} P_f(T_i) \tag{2}$$

Here, $P_f(T_i)$ is the probability density function of the input at i and is defined as given in equation (3).

$$P_f(T_i) = \frac{g_i}{g} \tag{3}$$

Here, g_i is the grayscale value of input pixel i and g is the total number of pixels in a block.

3.2 Removal of Pectoral Muscle

The pectoral muscle removal is an important stage before extracting ROI from the image because of the similarity in pixel intensity. Single Sided Edge Marking (SSEM) method is used in the proposed work for the removal of pectoral muscle.

3.2.1 Single Sided Edge Marking

The SSEM method is based on the geometric characteristics of mammograms and the intensity difference between muscle and breast tissue.

3.3 Feature Extraction

It is the most important step of image retrieval as the features contain relevant information that is required to distinguish between different classes.

3.3.1 GLCM Features

Grey Level Co-occurrence Matrix (GLCM) Method is a statistical way to examine texture by considering the spatial relation of pixels.

Contrast:

Contrast is a measure of the contrast intensity between a pixels and its neighbor over the entire image. It is calculated by using the equation (4).

$$Contrast = \sum_{u,v=0}^{N-1} G_{u,v}(u,v)^2 \tag{4}$$

Correlation:

Correlation is a measure of how a pixel is correlated to its neighbor over the entire image. It can be calculated by using the equation (5).

$$Correlation = \sum_{u,v=0}^{N-1} G_{u,v}\left(\frac{(u-\mu_u)(v-\mu_v)}{\sqrt{(\sigma_u^2)(\sigma_v^2)}} \right) \tag{5}$$

Energy:

Energy measures the repeated pairs. It is given by equation (6).

$$Energy = \sum_{u,v=0}^{N-1} G_{u,v}^2 \tag{6}$$

Homogeneity:

Homogeneity measures the actual distribution closeness of the entity in GLCM to GLCM diagonal. It can be calculated by the equation (7).

$$Homogeneity = \sum_{u,v=0}^{N-1} \frac{G_{u,v}}{1+(u-v)^2} \tag{7}$$

3.3.2 Gabor Features

Gabor Filter is an linear filter, that is widely being used to extract the texture features from the images for removal .The general function $F(u,v)$ of 2D Gabor filter is represented as given in equation (8).

$$F(u,v) = \frac{1}{2\pi\sigma_u\sigma_v} \exp\left[-\frac{1}{2}\left(\frac{u^2}{\sigma_u^2}+\frac{v^2}{\sigma_v}\right) + jw(u\cos\theta + y\sin\theta)\right] \tag{8}$$

Here θ is orientation, σ_u and σ_v define the Gaussian envelope is modulation frequency.

3.3.3 LBP Features

Local Binary Pattern (LBP) is the property of local imaging structures for performing the gray-scale and rotation invariant texture analysis.LBP operator is mathematically defined as given in (9).

$$LBP = \sum_{i=0}^{q-1} 2^n (I_n - I_c) \tag{9}$$

Here, I_c and I_n and are the gray scale values of the centre pixel and neighborhood pixel.

3.4 Classification

3.4.1 Optimized ANFIS

ANFIS classifier is trained and optimized uses an Improved Particle Swarm Optimization (IPSO) to enhance the classification accuracy.

3.4.2 Improved Particle Swarm Optimization:

The Particle Swarm Optimization (PSO) was designed on the basis on the social activities of a congregation of birds. In the proposed system, there is a hybridization that is done between two techniques to obtain better performance. The technique is called Improved PSO (IPSO). The pseudo code for IPSO is presented in Figure 2.

```
Begin
    For each and every particle
    Initialize position_i and velocity_i of the swarm.
    end for
    do
    for each and every particle
    Calculate fitness D  of the particle
        if D is more than P_best
    Set P_best = current fitness
        end if
        if P_best is better than G_best
    Set G_best = Best fitness value of all particles
        end if
        for each particle
    Find velocity_i, Update position_i
        for each particle
    Apply crossover & mutation
            end for
        end for
    end for
end
```

3.5 Similarity Assessment

In the proposed technique, similarity assessment is done after classification, so that it is enough to assess the similarity based on the classes such as normal, benign and malignant of classified query image. The Euclidean is the distance between feature value of the query image and the database images is represented here.

$$E = \sqrt{\sum_{i=}^{N} (q_i - d_i)^2} \tag{10}$$

Here, E is the Euclidean distance, q_i is the feature value of the query image and d_i is the feature value of the database images.

IV. RESULT AND DISCUSSION

4.1 Dataset Description

The proposed system uses the images of the mini-Mammography Image Analysis Society (MIAS) database of mammograms.

4.2 Performance Metrics

The performance metrics are calculated and evaluate to determine the efficiency of the proposed system. Some of the terms used in this analysis are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN)

TP denotes an normal images that are of correctly retrieved. FP denotes the images that are falsely classified. TN indicates the abnormal images that are correctly classified. FN stands for images that are falsely classified as abnormal mammogram images.

Sensitivity is a statistical measure of performance. Sensitivity is also called true positive rate. Accuracy determines how correctly the mammogram images are classified. The sensitivity, specificity as well as accuracy equations are mentioned below.

$$Sensitivit \quad y(i) = \frac{TP}{(TP + FN)} \qquad (11)$$

$$Specificit \quad y(i) = \frac{TN}{(TN + FP)} \qquad (12)$$

$$Accuracy \quad (i) = \frac{TP + TN}{(TP + TN + FP + FN)} \qquad (13)$$

Precision is a performance parameter that gives the fraction of normal images that are classified images among the classified instances. Recall is the fraction of relevant instances that have been classified over the total amount of classified instances.

$$\Pr ecision = \frac{TP}{(TP + FP)}$$

$$(14)$$

$$\text{Re} \, call = \frac{TP}{(TP + FN)} \qquad (15)$$

F-measure considers both recall as well as Precision to compute its value. Matthews Correlation Coefficient (MCC) is the correlation coefficient measure between the observed and the predicted classification.

$$F\_measure = \frac{2*(\Pr ecision * \text{Re} \, call)}{(\Pr ecision + \text{Re} \, call)} \qquad (16)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TP + FN)}} \qquad (17)$$

The comparative analysis of different classifiers can be done in terms of Negative Predictive Value (NPV), False Positive Rate (FPR), and False Negative Rate (FNR). The NPV is the probability that the mammogram images are abnormal and are truly abnormal images. This abnormality indicates whether the image is benign or malignant. The FPR and FNR are also evaluated. These are done using the following equations.

$$FPR = \frac{FP}{FP + TN} \qquad (18)$$

$$FNR = \frac{FN}{FN + TP} \qquad (19)$$

4.3 Comparative Analysis

4.3.1 Comparative Analysis of Mammogram Image Retrieval based on Classification
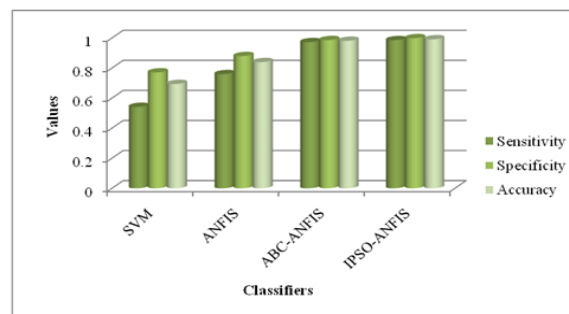

Figure 4: Comparison of the existing and the proposed classifiers with respect to Sensitivity, Specificity Accuracy

Figure 4 analyzes function of IPSO-ANFIS with the other existing techniques. The proposed IPSO-ANFIS shows the greatest performance. But, among ABC-ANFIS and the proposed one, the proposed IPSO-ANFIS shows superior performance.
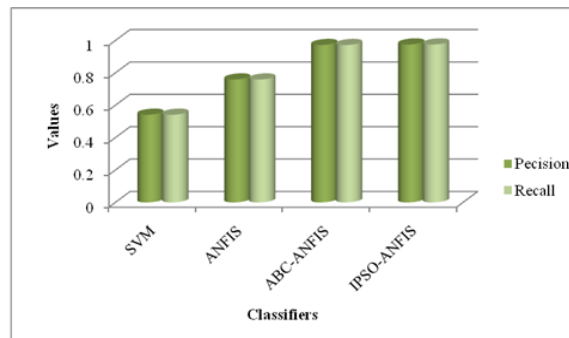

Figure 5: Comparison of the existing and the proposed classifiers in terms of Precision and Recall

Figure 5 analyzes the function of the IPSO-ANFIS with the other existing techniques. On considering Precision, SVM shows very poor performance and the performance of ANFIS is better when compared to SVM.

Classifiers	F-Measure	MCC
SVM	0.54167	0.3125
ANFIS	0.75926	0.63889
ABC-ANFIS	0.97222	0.95833
Proposed IPSO-ANFIS	0.97542	0.98541

Table 1: Comparison of the existing and the proposed classifiers in terms of F-Measure and MCC

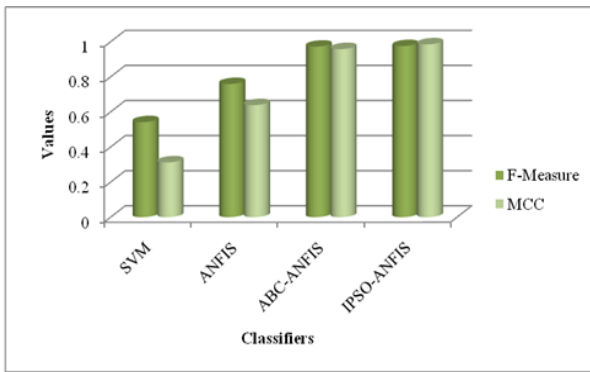Table 1 compares the performance of various methods.

Figure 6: Comparison of the existing and the proposed classifiers in terms of F-Measure and MCC

Figure 6 analyzes the function of the proposed IPSO-ANFIS with the other existing techniques. It can be seen that there is an improvement in both the F-Measure and the MCC when the optimization technique is combined with the ANFIS classifier.
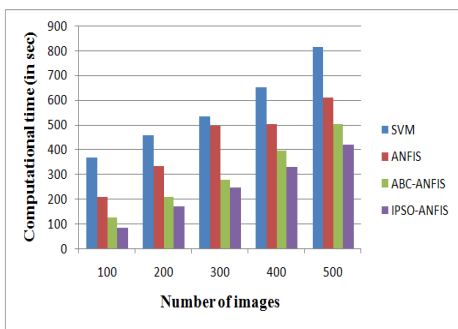

Figure 7: illustrate the computational time of the proposed classifier with the existing classifiers

Figure 7 shows computational time comparison of the proposed IPSO-ANFIS classifier with the existing classifiers. Hence it proves the proposed IPSO-ANFIS classifier has better performance when compared with the other existing classifiers.

V. CONCLUSION

This paper puts forward an efficient Content-Based Mammogram Image Retrieval technique by means of an IPSO Optimized Classifier. On evaluating all the comparisons that are discussed in this paper, it can be said that the proposed technique is superior to all other compared existing techniques. All these show that the proposed system is highly efficient in retrieving the mammogram images.

VI. ACKNOWLEGEMENT

VII. REFERENCE

1. Kanchan Lata Kashyap, Manish Kumar Bajpai, and Pritee Khanna, "An efficient algorithm for mass detection and shape analysis of different masses present in digital mammograms", Multimedia Tools and Applications, Vol. 77, No. 8, pp. 9249-9269, 2018.
2. Vibhav Prakash Singh, Ashim Gupta, Shubham Singh, and Rajeev Srivastava, "An efficient content based image retrieval for normal and abnormal mammograms", In Electrical Computer and Electronics (UPCON), 2015 IEEE UP Section Conference on, IEEE, pp. 1-6, 2015.
3. Lazaros Tsochatzidis, Konstantinos Zagoris, Nikolaos Arikidis, Anna Karahaliou, Lena Costaridou, and Ioannis Pratikakis, "Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach", Pattern Recognition, vol. 71, pp. 106-117, 2017.
4. Sami Dhahbi, Walid Barhoumi, and Ezzeddine Zagrouba, "Content-Based Mammogram Retrieval Using Mixed Kernel PCA and Curvelet Transform", In International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, Cham, pp. 582-590, 2016.
5. Marcos Vinicius Naves Bedo, Davi Pereira dos Santos, Marcelo Ponciano-Silva, Paulo Mazzoncini de Azevedo-Marques, and Caetano Traina, "Endowing a content-based medical image retrieval system with perceptual similarity using ensemble strategy", Journal of Digital Imaging, vol. 29, no. 1, pp. 22-37, 2016.
6. Fradj Ben Lamine, Karim Kalti, and Lotfi Ben Romdhane, "A content-based digital mammography retrieval using inexact graph matching", In Image Processing, Applications and Systems Conference (IPAS), 2014 First International, IEEE, pp. 1-4, 2014.
7. Shobha Jose, and D. Abraham Chandy, "Content based mammogram retrieval using biorthogonal wavelet filters in DDSM database", In Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on, IEEE, pp. 1-6, 2014.
8. Devang Kulshreshtha, Vibhav Prakash Singh, Ayush Shrivastava, Arpit Chaudhary, and Rajeev Srivastava, "Content-based mammogram retrieval using k-means clustering and local binary pattern", In Image, Vision and Computing (ICIVC), 2017 2nd International Conference on, IEEE, pp. 634-638, 2017.
9. Thenkalvi Boomilingam, and Murugavalli Subramaniam, "An efficient retrieval using edge GLCM and association rule mining guided IPSO based artificial neural network", Multimedia Tools and Applications, vol. 76, no. 20, pp. 21729-21747, 2017.
10. Vaidehi, K., and T. S. Subashini, "Automatic classification and retrieval of mammographic tissue density using texture features", In Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on, IEEE, pp. 1-6, 2015.
11. Juan Wang, and Yongyi Yang, "Feature saliency analysis for perceptual similarity of clustered microcalcifications", In ICIP, pp. 775-778, 2015.
12. Abraham Chandy, D., A. Hepzibah Christinal, Alwyn John Theodore, and S. Easter Selvan, "Neighbourhood search feature selection method for content-based mammogram retrieval", Medical & Biological Engineering & Computing, vol. 55, no. 3, pp. 493-505, 2017.
13. Ramzi Chaieb, and Karim Kalti, "Feature subset selection for classification of malignant and benign breast masses in digital mammography", Pattern Analysis and Applications, pp. 1-27, 2018.
14. Menglin Jiang, Shaoting Zhang, Hongsheng Li, and Dimitris N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval", IEEE Transactions on Biomedical Engineering, vol. 62, no. 2, pp. 783-792, 2015.
15. Liliana Losurdo, Annarita Fanizzi, Teresa MA Basile, Roberto Bellotti, Ubaldo Bottigli, Rosalba Dentamaro,

Vittorio Didonna et al, "A Combined Approach of Multiscale Texture Analysis and Interest Point/Corner Detectors for Microcalcifications Diagnosis", In International Conference on Bioinformatics and Biomedical Engineering, Springer, Cham, pp. 302-313, 2018.

16. Vibhav Prakash Singh, and Rajeev Srivastava, "Automated and effective content-based mammogram retrieval using wavelet based CS-LBP feature and self-organizing map", Biocybernetics and Biomedical Engineering, Vol. 38, No. 1, pp. 90-105, 2018.

17. Mugahed A. Al-antari, Mohammed A. Al-masni, Sung-Un Park, JunHyeok Park, Mohamed K. Metwally, Yasser M. Kadah, Seung-Moo Han, and Tae-Seong Kim, "An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network", Journal of Medical and Biological Engineering, Vol. 38, No. 3, pp. 443-456, 2018.

18. Syed Jamal Safdar Gardezi, Ibrahima Faye, Jose M. Sanchez Bornot, Nidal Kamel, and Mohammad Hussain, "Mammogram classification using dynamic time warping", Multimedia Tools and Applications, Vol. 77, No. 3, pp. 3941-3962, 2018.

19. Birmohan Singh, and Manpreet Kaur, "An approach for classification of malignant and benign microcalcification clusters", S?dhan?, Vol. 43, No. 3, pp. 39, 2018.

20. Ardalan Ghasemzadeh, Saeed Sarbazi Azad, and Elham Esmaeili, "Breast cancer detection based on Gabor-wavelet transform and machine learning methods", International Journal of Machine Learning and Cybernetics, pp. 1-10, 2018.

21. Chisako Muramatsu, "Overview on subjective similarity of images for content-based medical image retrieval", Radiological Physics and Technology, Vol. 11, No. 2, pp. 109-124, 2018.

22. Saravanan, M, Sukanya, S, Image based password authentication system for banks, International Conference on Information Communication and Embedded Systems, ICICES 2017,2017

23. Rajesh Kannan, Dr.Meera Gandhi (2012), "Dynamic key exchange method for image encryption", International Journal of Computer Application, issue 2, vol. 1, pp. 99-105, February 2012.

24. Nirmalrani, V., Saravanan, P., Sakthivel, P., An extended XACML model to secure biological web services using access control policies, Research Journal of Pharmaceutical, Biological and Chemical Sciences, 2016.