Proceedings of the 11th INDIACom; INDIACom-2017; IEEE Conference ID: 40353
2017 4th International Conference on "Computing for Sustainable Global Development", 01st - 03rd March, 2017
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

Classification of Breast Lesions Using Texture Ratio Vector Technique

Sahil Bhusri
Lovely Professional University
Phagwara, Punjab, India
Email ID: sahilbhusri1@gmail.com

Shruti Jain
Jaypee University of Information Technology, Solan, India
Email ID: shruti.jain@juit.ac.in

Abstract— Breast cancer is the most common cancer in women. It arises due to the uncontrolled growth of cells in the breast. The area suffering from damage is known as lesion and is classified into two categories i.e. Benign and Malignant. This paper classifies the breast lesions using a ratio texture feature obtained from the texture features calculated inside the lesion (IAI) and the texture feature calculated from the upper side of the lesion (UAI). Statistical texture features like EDGE, SFM, NGTDM, FOS, GLCM, GLRLM and GLDS were calculated. SVM classifier is used to classify the lesions on the basis of ratio texture feature. The texture features calculated from IAI gains an overall accuracy of 62.2 % with NGTDM texture feature where as an overall accuracy of 82.2 % is achieved in UAI using the GLCM texture feature. However an overall accuracy of 86.6% is yielded with the FOS ratio texture vector having individual accuracies of 82% and 92.2% for benign and malignant class respectively.

Keywords — Breast cancer, Breast Lesion classification, Ratio Texture Vector, Statistical Texture Features, Benign, Malignant, Ultrasound

I. INTRODUCTION

A million number of cells combines together to make a tissue or organ . These cells are changed or reconstructed at a regular interval. Whenever there is uncontrolled growth of cells starts in body, it leads to the beginning of the disease of cancer. Breast cancer is a form of cancer that originates in the breast tissues. It starts in the lobules or in the duct of breast that carries the milk [1]. Early diagnosis is having the privilege to save the patient from unwanted result of the disease. A variety of breast screening and detection techniques are available these days, any change in breast size, lump or any deformity in breast can be detected with the mammography, breast ultrasound and breast MRI. These techniques don't only help in early detection of disease but also diminishes the chances of unwanted biopsies. There are two types of biopsies available i.e. Needle biopsy and Surgical biopsy. [2].So in order to reduce the chances of biopsies, these screening methods are used as second opinion to doctor in making the decision. The imaging systems i.e. X-Ray, Ultrasound and MRI provide the internal view of the breast. In X- ray, a radiation is passed through the body and when it strikes to the photographic plate, an image is produced. In Ultrasound, a high frequency sound wave is passed to the internal organs, the echo that returns back creates an image where as in MRI, a strong magnet and radio waves pulses are used to create an image. In young patients, where the tissues of breast are very dense, X- ray has very low sensitivity and it doesn't disclose the soft tissues, where as MRI is very expensive and does require more time for the perfect scan . Therefore ultrasound imaging is preferred due to its better penetration, real time display and low cost [3]. BIRADS system is used to standardize the reporting of breast cancer and used as a quality assurance tool for ultrasound images [2]. Ultrasonography distinguishes between the non invasive (Benign) and invasive (Malignant) form of cancer .It also helps to distinguish between the cyst and the solid mass present. The most commonly occurring benign case is Fibroadenoma, it occurs in 95% cases of non invasive cancer and the mast commonly occurring malignant cancer is Carcinoma. Different CAD tools have been designed earlier to distinguish the breast lesions. It has been investigated earlier that an overall accuracy of 84% is achieved using the morphological features to distinguish the lesions [4] and an overall accuracy of 85% [5] is obtained using the difference of statistical feature vector between inner ROI and the outer ROI of lesion.In this paper, the breast lesions are characterized in to benign and malignant lesions using the computation of texture ratio feature vector. The texture ratio feature vector is created using the ratio of texture feature present inside the lesion (*IAI*) and the texture features present outside the lesion i.e. on the upper side of the lesion (*UAI)*.

II. METHODOLOGY

CAD system designed for classification of malignant and benign lesion follows a sequence as shown in Figure 1

A. Input Ultrasound Database Images

A standard set of ultrasound images data is taken online from [6] for the benign and malignant cases. The dataset contains a total of 117 cases having 45 cases of benign class and 72 cases of malignant class. The sample of benign and malignant is displayed in Fig 2. The data set is divided into further two subcategories i.e. Training dataset and Testing Dataset. (Note : cases of lesions during biopsy implementation are discarded.)

B. ROI (Region of Interest) Extraction Module

The irregularity in the ultrasound is marked with help of an experienced radiologist. The infected area in image is marked and segmented with help of software *Image J* [7]. A variable size rectangular region of interest is taken from inside the lesion (*IAI*) and a rectangular region of size 32 × 32 is taken from upper side of the lesion (*UAI*) .i.e. from outside the lesion. The major concern of taking IAI and UAI is because of taking the ratio of texture features present in the damaged area and texture features of area surrounding that damaged area. The samples of IAI and UAI are represented in Fig 3 and Fig 4.

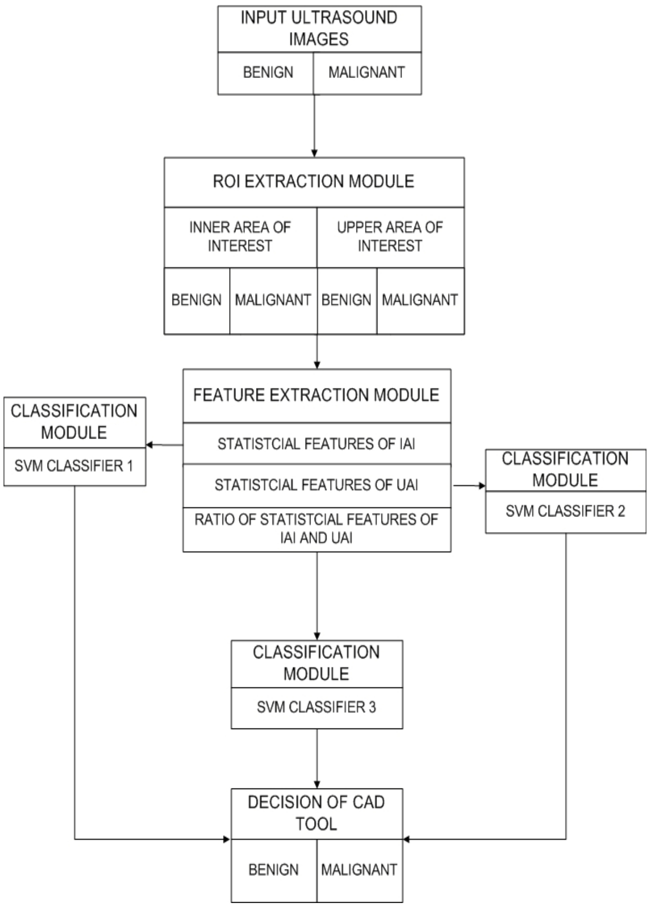

Fig. 1. Overview of system

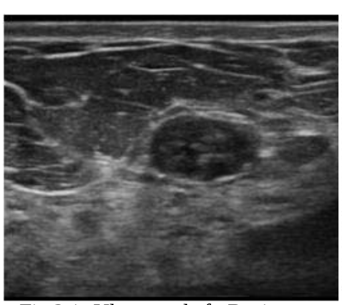
Fig 2.1: Ultrasound of Benign Case

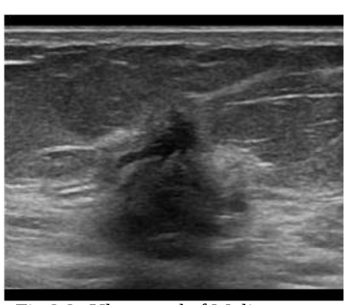
Fig 2.2: Ultrasound of Malignant Case

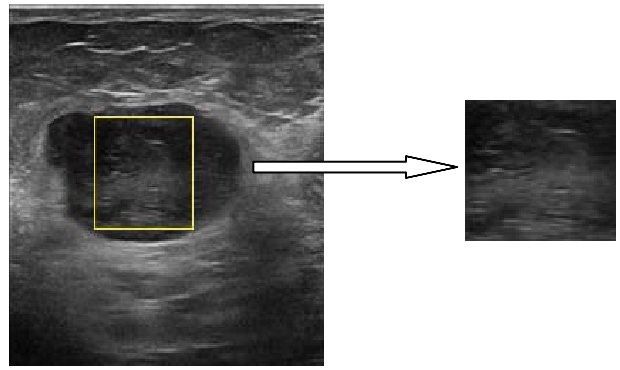
Fig. 3.1 Fig. 3.2

Figure 3.1 : IAI marked Ultrasound
Figure 3.2 : Extraction of IAI from ultrasound

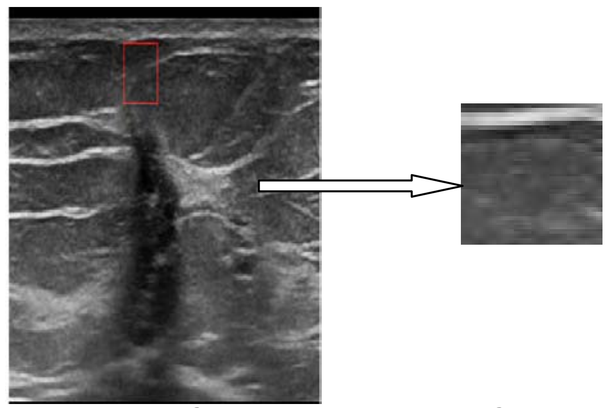
Fig.3.1 Fig. 3.2

Figure 3.1 : UAI marked Ultrasound
Figure 3.2 : Extraction of UAI from ultrasound

C.. Feature extraction Module

The area suffering from disease is known as lesion. Breast lesions can be characterized by analyzing the shape, geometry, spiculations and microlobulation [8], but in most of the cases there is not enough information present, therefore feature extraction techniques are used to characterize the breast lesions . Feature extraction module offers different methods to extract the features of the lesion. It includes texture feature extraction techniques and the morphological feature extraction techniques. The morphological feature extraction techniques uses the shape of lesion to characterize where as the texture feature extraction module extracts features using the properties of surface. Texture features (TF) can be extracted with 1) Signal Processing Methods 2) Statistical Methods and 3) Transform Domain Methods. This paper includes the characterization of lesions using the Statistical Texture Feature extraction techniques. Statistical features are computed from IAI and UAI. A ratio vector of these texture features is created to characterize the lesion [9].

1) Statistical Methods

An ultrasound image has pixels of various gray intensities. Based on these gray level intensities, the statistical features are computed using different order statistical methods.

First order statistical methods computes average gray level, entropy, roughness and uniformity using the histogram moments to describe the texture[10].

Second order statistical method uses Gray Level Co-occurrence Matrix (GLCM) to compute the texture features. It uses the frequent combinations of gray level intensities of same gray intensity levels present in different directions say θ = 0°, 45°, 90°, 135°. A set of 13 GLCM TF were computed [11-12].

Higher order statistical features use Gray Level Run Length Matrix (GLRLM) to compute the TF. It uses the persistent combinations of gray level intensities present at a relative distance at each other. The consecutive set of gray level pixels constitute gray level run and no of times run occurs constitute run length. A set of 11 GLRLM are computed [13-14].

Other statistical features compute Gray Level difference statistics (GLDS), Statistical Feature Matrix (SFM), Neighborhood Gray Tone Difference Matrix (NGTDM) and Edge Features. GLDS calculates energy, entropy, contrast, mean and homogeneity depending upon the pixel pairs of different gray level present at different distances from each other. SFM calculates coarseness, contrast, periodicity and roughness of pixels at different distances within an ultrasound image. NGTDM consider the distance present in between the pixels and used to calculate the contrast, coarseness, complexity and busyness where as EDGE features extracts information present at the edge of the images. These values can be computed using the gradient value. The value of gradient will be high if more information is present else it will be low if no information is present[15-16].

D. *Classification Module*

The process of grouping the testing class samples into the different classes is known as classification process. If a set of class is already defined for the training set then the classification is supervised classification and if the training set is not available then it is unsupervised classification. The SVM classifier is a form of supervised classifier used for the classification. It is working on basis of statistical theory and used for the linear and non linear classification. It creates a hyper plane with the help of training set available and good separation is achieved instinctively. In non linear classification problems, the input data is mapped in to the kernel functions in which the data is mapped from input space to the higher dimensional feature space. For the classification task, Gaussian radial basis function kernel's are used. LibSVM library has been used for the implementation of SVM classifier [17-45].

III. Results and Discussion

In this paper, a total of 117 cases are considered for classification purpose. These cases have 45 cases of benign class and 72 cases malignant class. The ratio vector is created for both the regions taken between IAI and UAI .Different experiments have been performed in this paper are tabulated in Table 1 where as the results are shown in Table 2 to Table 4.

TABLE I. List of experiments performed to classify the breast lesions

Exp 1	To obtain classification performance of texture features obtained from IAI
Exp 2	To obtain classification performance of texture features obtained from UAI
Exp 3	To obtain classification performance of texture features obtained from ratio of statistical features between IAI and UAI

The performance of various feature vectors (*FVS*) obtained from IAI using SVM classifier is tabulated in Table 2

TABLE II. Classification performance of various FVS obtained from IAI

TABLE III.

FVS	CM			ICA	OCA
		B	M		
EDGE	B	12	8	60.0 %	62.2%
	M	9	16	64.0 %	
FOS	B	9	11	45 .0%	53.3%
	M	10	15	60.0%	
GLCM	B	4	16	20.0%	46.6%
	M	8	17	68.0 %	
GLDS	B	8	12	40.0 %	55.7%
	M	8	17	68.0 %	
GLRLM	B	0	20	0.0 %	44.4%
	M	5	20	80.0 %	
NGTDM	B	13	7	65.0 %	62.2%
	M	10	17	68.0 %	
SFM	B	10	10	50 %	55.7
	M	10	15	60 %	

Note: FVS: Feature Vectors, CM :Confusion matrix , OCA : Over all classification accuracy , B: benign class , M: Malignant Class , ICA: Individual class accuracy , FOS : First order statistics , GLCM : Gray length co-occurrence matrix , GLRLM : Gray level run length matrix , GLDS: Gray level difference statistics , NGTDM : Neighborhood gray tone difference matrix., SFM : Statistical feature matrix. *The feature vector having best OCA has been shaded with gray background.*

From Table 2 it is noticed that the NGTDM feature has highest overall accuracy of 62.2 % in IAI with individual accuracies of 65.0 % and 68.0 % for benign and malignant classes respectively.

TABLE IV. CLASSIFICATION PERFORMANCE OF VARIOUS FVS OBTAINED FROM UAI

FVS	CM			ICA	OCA
		B	M		
EDGE	B	15	5	75.0 %	62.2%
	M	5	20	80.0 %	
FOS	B	15	5	75.0 %	75.5%
	M	6	19	76.0 %	
GLCM	B	15	5	75.0 %	82.2%
	M	3	22	88.0 %	
GLDS	B	9	11	45.0 %	66.6%
	M	4	21	84.0 %	
GLRLM	B	0	20	0.0 %	55.5%
	M	0	25	100 %	
NGTDM	B	13	7	65.0 %	71.7%
	M	6	19	76.0 %	
SFM	B	15	5	75.0 %	62.2%
	M	12	13	52.0 %	

Note : *The feature vector having best OCA has been shaded with gray background*

From Table 3 it is noticed that in UAI, GLCM feature is providing the highest overall accuracy of 82.2 % in with the highest individual accuracies of 75 % and 88 % for benign and malignant classes respectively.

TABLE V. CLASSIFICATION PERFORMANCE OF VARIOUS FVS OBTAINED WITH RATIO VECTOR OF IAI AND UAI

FVS	CM			ICA	OCA
		B	M		
EDGE	B	4	16	20.0 %	46.6%
	M	8	17	68.0 %	
FOS	B	16	4	80.0 %	86.6%
	M	2	23	92.2 %	
GLCM	B	9	11	45.0 %	64.4%
	M	5	20	80.0 %	
GLDS	B	12	8	60.0 %	55.5%
	M	12	13	52.2 %	
GLRLM	B	0	20	0.0 %	44.4%
	M	5	20	75.0%	
NGTDM	B	3	17	15.2 %	35.5%
	M	12	13	52.2 %	
SFM	B	4	16	20.0 %	40%
	M	11	14	56.0 %	

Note: *The feature vector having best OCA has been shaded with gray background.*

From Table 4, It has been noticed that in classification with the ratio vector of IAI and UAI, FOS has achieved the highest accuracy of 86.6 % with individual accuracies of 80 % and 92.2 % for benign and malignant classes respectively.

IV. CONCLUSION

This paper proposes a CAD system to help the radiologist as second opinion in classification of breast lesions. The experiments carried out in this paper present that the FOS texture feature ratio vector of UAI and IAI gains the maximum overall accuracy of 86.6 % in classification of the lesions and gains the individual accuracy of 80 % for benign class and individual accuracy of 92.2 % for the malignant class.

V. ACKNOWLEDGMENT

The authors want to acknowledge Dr Shruti Thakur, IGMC, Shimla, Himachal Pradesh, India and Dr Jitendra Virmani, Sr. Technical Officer, CSIO, Chandigarh for their constant support and valuable suggestions in carrying out this research work

REFERENCES

[1] *Cancer Research UK* [online] ,2016 [cited 2016 ,March] http://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancer-starts
[2] Breast Biopsies, Susan G. Komen [online],2016[citied 2016, October] http://ww5.komen.org/BreastCancer/Biopsies.html
[3] Bad Back Store, Medical Image Testing [online], 2016 [citied 2016, October] . http://www.badbackstore.com/news_a/medical_Imaging_Tests_difference_Between_X_Ray_a/126.html
[4] A. V. Alvarenga, A. F. C. Infantosi, W. C. A. Pereira, and C. M. Azevedo, "Assessing the performance of morphological parameters in distinguishing breast tumors on ultrasound images," *Medical engineering & physics*, vol. 32, no. 1, pp. 49–56, 2010
[5] S. Bhusri , S. Jain and J .Virmani , " Classification of Breast Lesions using the difference of statistical features" Using Image Processing Techniques Research Journal of Pharmaceutical, Biological and Chemical Sciences (RJBPS) , vol.7 , issue 4, pp. 1365-1372 , 2016
[6] *Ultrasound Cases.info* , [online] ,2015 http://www.ultrasoundcases.info/category.aspx?cat=67
[7] *Image Processing and Analysis in JAVA* ,Image J 1.49 vesion1.6.024 *http://imagej.nih.gov/ij/download/win32/ij149-jre6-64.zip*
[8] MO Obajimi, OO Akute, AO Afolabi, AA Adenipekun, AO Oluwasola, EEU Akang, RU Joel, ATS Adeniji-Sofoluwe, O Funmi, G Newstead, R Schmitt, C Sennet ,''BI-RADS Lexicon : An Urgent Call for the standardization of Breast Ultrasound in Nigeria" *Annals of Ibadan Postgraduate Medicine* Vol. 3 issue 1, pp. 82-88 , 2005
[9] S.Bhusri, S.Jain , J.Virmani, "Breast Lesion Classification using amalgamation of Morphological and Texture Features", International Journal of Pharma and BioSciences, vol 7, no.2 ,pp : (B) 617 – 624, 2016
[10] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal,"A rapid approach for prediction of liver cirrhosis based on first order statistics", in Proceedings of the IEEE International Conference on Multimedia, Signal Processing and Communication Technologies, IMPACT-2011, Aligarh, India, pp. 212-215, 2011.

[11] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal , "Prediction of cirrhosis based on singular value decomposition of gray level co-occurrence matrix and a neural network classifier", in *Proceedings of Development in E-systems Engineering, DeSE*, Dubai, pp. 146-151, 2011.

[12] P. Mohanaiah, P. Sathyanarayanam and L. Gurukumar, "Image texture feature extraction using GLCM approach", *International Journal of Scientific and Research Publications*, vol. 3, pp. 1-5, 2013

[13] D.H. Xu, A.S. Kurani, J.D. Furst and D.S. Raicu, "Run-length encoding for volumetric texture", *Heart*, 2004 ; vol. 27: 25-30.

[14] F. Albregtsen., "Statistical texture measures computed from gray level run length matrices", *Image*,; vol. 1 ,pp 3-8, 1995.

[15] J.S. Weszka, C.R. Dyer and A. Rosenfeld, "A comparative study of texture measures for terrain classification", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6, pp. 269-285, 1976.

[16] J.K. Kim and H.W. Park, "Statistical textural features for detection of microcalcifications in digitized mammograms", *IEEE Transactions on Medical Imaging*, vol. 18, 1999.

[17] C.C. Chang and C. J. Lin, "LIBSVM, A library of support vector machines".

[18] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "Neural network ensemble based CAD system for focal liver lesions from B-mode ultrasound", *Journal of Digital Imaging*, vol. 27, pp. 520-537, 2014.

.

[19] Kriti, J. Virmani, N. Dey and V. Kumar, "PCA-PNN and PCA-SVM based CAD systems for breast density classification", in A.E. Hassanien et al. (eds.) *Applications of Intelligent Optimization in Biology and Medicine: Current Trends and Open Problems*, vol. 96, pp. 159-180, 2015.

[20] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "SVM based characterization of liver cirrhosis by singular value decomposition of GLCM matrix", *International Journal of Artificial Intelligence and Soft Computing*, vol. 3, pp. 276-296, 2013.

[21] A.E. Hassanien, N.E. Bendary, M. Kudelka and V. Snasel, "Breast cancer detection and classification using support vector machines and pulse coupled neural network", in *Proceedings of 3rd International Conference on Intelligent Human Computer Interaction IHCI 2011,* Prague, Czech Republic, 2011, pp. 269-279.

[22] S Jain, PK Naik, SV Bhooshan. "Nonlinear Modeling of cell survival/ death using artificial neural network"., International Conference on Computational Intelligence and Communication Networks (CICN2011), Gwalior, India , pp 565-568, Oct 07-09, 2011

[23] S. Jain , P.K. Naik , "System Modeling of cell survival and cell death : A deterministic model using Fuzzy System", International Journal of Pharma and BioSciences(IJPBS); vol. 3, issue 4, pp. 358-373 ,2012

[24] A.T.Azar and S.A. El-Said, "Performance analysis of support vector machine classifiers in breast cancer mammography recognition", *Neural Computing and Applications*, vol. 24, pp. 1163-1177, 2014

[25] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "PCA-SVM based CAD system for focal liver lesion using B-mode ultrasound images", *Defence Science Journal*, vol. 63, pp. 478-486, 2013.

[26] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors", *Journal of Digital Imaging*, vol. 26, no. 3, pp. 530-543, 2012.

[27] S Jain , "Compedium model using frequency / cumulative distribution function for receptors of survival proteins: Epidermal growth factor and insulin", 6(4), 101-110 : Dec 2016.

[28] S Rana, S Jain, J Virmani, "SVM-Based Characterization of Focal Kidney Lesions from B-Mode Ultrasound Images ",Research Journal of Pharmaceutical, Biological and Chemical Sciences (RJPBCS),7(4), 837-846: July- Aug, 2016.

[29] A Dhiman, A Singh, S Dubey, S Jain, "Design of Lead II ECG Waveform and Classification Performance for Morphological features using Different Classifiers on Lead II ", *Research Journal of Pharmaceutical, Biological and Chemical Sciences (RJPBCS)*,7(4), 1226-1231: July-Aug 2016.

[30] S. Bhusri , S. Jain, J. Virmani, "Classification of Breast Lesions based on Laws'Feature Extraction Techniques ", March 16th- 18th, 3rd 2016 International Conference on Computing for Sustainable Global Development, BVICAM, New Delhi. , pp 1700-1704, 2016

[31] Shruti Jain, " Parametric and Non Parametric Distribution Analysis of AkT for Cell Survival/Death" , International Journal of Artificial Intelligence and Soft Computing, 2016.

[32] S Jain, "Regression analysis on different mitogenic pathways", Network Biology, 6(2), 40-46 : June 2016.

[33] S Jain, "*Mathematical Analysis using Frequency and Cumulative Distribution functions for Mitogenic Pathway", Research Journal of Pharmaceutical, Biological and Chemical Sciences (RJPBCS)*,7(3), 262-272: May - Jun 2016

[34] S Jain, D. S. Chauhan, "Mathematical Analysis of Receptors For Survival Proteins", International Journal of Pharma and Bio Sciences (IJPBS),6(3), 164-176 : 2015.

[35] S Jain, D. S. Chauhan, "Implementation of fuzzy system using different voltages of OTA for JNK pathway leading to cell survival/ death", Network Biology, 5(2), 62-70 : 2015.

[36] S Jain, "Design and Simulation of Fuzzy Implication Function of Fuzzy System Using Two Stage CMOS Operational Amplifier", International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 150-155, 7(2): Dec 2013- Feb 2014.

[37] S Jain, "Design and simulation of fuzzy membership functions for the fuzzification module of fuzzy system using operational amplifier", International Journal of Systems, Control and Communications (IJSCC), 69-83, 6(1):2014.

[38] S Jain, P. K. Naik, "System Modeling of cell survival and cell death : A deterministic model using Fuzzy System", International Journal of Pharma and BioSciences (IJPBS),358-373, 3(4): 2012

[39] **S** Jain, S. V. Bhooshan, P. K. Naik, "Compendium Model of AkT for Cell Survival/Death and its Equivalent BioCircuit", International Journal of Soft Computing and Engineering (IJSCE), 91-97, 2(3): July 2012 (ISSN : 2231-2307)

[40] S Jain, Sunil V. Bhooshan, Pradeep K. Naik, "Mathematical modeling deciphering balance between cell survival and cell death using Tumor Necrosis Factor α", *Research Journal of Pharmaceutical, Biological and Chemical Sciences (RJPBCS)*, 574-583, 2(3): , 2011, (ISSN : 0975-8585)

[41] S.Jain, S. V. Bhooshan, P. K. Naik, "Model of Mitogen Activated Protein Kinases for Cell Survival/Death and its Equivalent Bio-Circuit", Current Research Journal of Biological Sciences (CRJBS), 2(1):59-71, 2010.

[42] S Jain , "Mathematical Analysis and Probability Density Function of FKHR pathway for Cell Survival /Death" , Aug 1-2, 2015, pp 84-93, Control System and Power Electronics – CSPE 2015, Banglore 2015.

[43] **S** Jain, D.S.Chauhan, "Linear and Non Linear Modeling of Protein Kinase B/ AkT", July 3-4 2015, pp 81-88 , International Conference on Information and Communication Technology for Sustainable Development (ICT4SD - 2015), Ahmedabad, India.

[44] S Jain , P.K. Naik, S. V. Bhooshan, "A System Model for Cell Death using Fuzzy Logic and SPICE", February 26-27, 2010, pp 756-760, ICWET 2010, International Conference and Workshop on Emerging Trends in Technology, Thakur College of Engineering & Technology,Mumbai, India.

[45] S Rana , S Jain, J Virmani, "Classification of Kidney Lesions using Gabor Wavelet Texture Features", March 16th- 18th, 2016, pp 2528-2532, 10th INDIACom: 3rd 2016 International Conference on Computing for Sustainable Global Development, BVICAM, New Delhi.