International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Mammogram Classification using Curvelet Coefficients and Gray Level Co-Occurrence Matrix for Detection of Breast Cancer

Ranjit Biswas, Sudipta Roy, Abhijit Biswas

Abstract: ***Computer-aided diagnosis system plays an important role in diagnosis and detection of breast cancer. In computer-aided diagnosis, feature extraction is one of the important steps. In this paper, we have proposed a method based on curvelet transform to classify mammogram images as normal -abnormal, benign and malignant. The feature vector is computed from the approximation coefficients. Directional energy is also calculated for all sub-bands. To select the efficient feature we used t-test and f-test methods. The selected feature is applied to Artificial Neural Network (ANN) classifier for classification. The effectiveness of the proposed method has been tested on MIAS database. The performance measures are computed with respect to normal vs. abnormal and benign vs. malignant for using approximation subband and energy feature of all curvelet coefficients. The highest classification accuracy of 95.34% is achieved for normal vs. abnormal and 80.86% is achieved for benign vs malignant class using energy feature of all curvelet coefficients.***

Keywords**:** ***Mammogram, ROI, Curvelet Transform, GLCM, Artificial Neural Network (ANN).***

I. INTRODUCTION

Breast cancer is one of the most prominent cancers worldwide. According to the findings of Globocan project, breast cancer contributes to a staggering 25% of the all cancers worldwide. India also witnesses a significant number of breast cancer cases every year and the data available shows the morbidity rate is almost equal to 50% of all the cases detected [1]. The key to improve survival rate amongst breast cancer patients lies in early detection of this terminal disease. Mammography is regarded as one of the most efficient ways of detecting breast cancer [2]. Mammography images are scan images of breast, these images aide radiologists in deciding whether a patient may be advised to undergo biopsy or not. A computer aided diagnostic system may prove very helpful in classifying a benign tumour from a malignant one. Also a CAD based system has a very low error rate in classifying abnormal types [3,4].

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Ranjit Biswas*, Department of Information Technology, Ramkrishna Mahavidyalaya, Kailashahar, India, Email: ranjitit1984@gmail.com
Sudipta Roy, Department of Computer Science & Engineering, Assam University, Silchar, India. Email: sudipta.it@gmail.com
Abhijit Biswas, Department of Computer Science & Engineering, Assam University, Silchar, India. Email: abhi.021983@gmail.com

Many researchers have contributing CAD based systems for digital mammogram classification. In [5] authors have described a classification of mammogram on the basis of wavelet decomposition and gray level image structure features. Wei et al. has achieved 96% area under curve by employing multi-resolution texture features for mammogram classification [6]. In [7] a set of wavelet based statistical features have been used, these features are subjected to binary tree classifier to obtain 84.2% accuracy. Daubechies wavelets based classification technique also obtains an accuracy rate of over 87% in mammogram classification [8]. SVM based classifiers were also demonstrated on mammogram features which were extracted using spherical wavelet transform (SWT) and achieved over 90% accuracy in classifying both malignant and benign tumours in a mammogram image [9].

II. PROPOSED METHOD

The proposed mammographic image classification method consists of four prime modules, such as region of interest (ROI) extraction, feature extraction, feature selection, and classification. A multi-resolution and multi-orientation feature extractor, called curvelet transform is employed on mammographic ROI images to extract curve features, where ROI image is decomposed into several sub-images. Once the feature matrix is generated, the feature selection procedure is performed using two-sample t-test and f-test approach to reduce the dimension of the features. In the feature selection algorithm; effective and significant features are selected and are provided to the neural network for the classification of mammograms as normal or abnormal and benign or malignant classes. The detailed block diagram of the proposed method is shown in Fig. 1 and the different steps are explained in the following section.

A. Region of interest (ROI) extraction process

The original mammogram images are of 1024x1012 pixels, which have different types of noises, artifacts in their background, pectoral muscles, etc. All of these are unsuitable for feature extraction and classification. Hence a cropping operation have been applied on mammographic images to extract the abnormal area of the mammogram images and remove the unwanted portions of the images, which is named as the region of interest (ROI). The MIAS database [10] gives all the details by the expert radiologist about each mammogram image, viz., size in pixels,

the character of

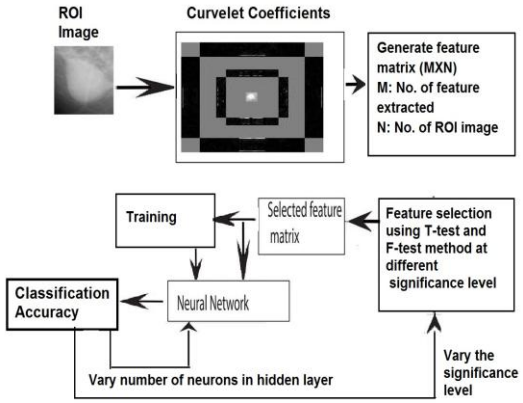


Fig. 1.Block diagram of the proposed system.

background tissue, class of abnormality, Xc and Yc coordinate value of the center of abnormality, and 'r' radius of a circle containing the abnormal area. The ROIs extraction process is performed by referring to the center of the abnormal area which is the center of ROI and taking the approximate radius (in pixels) of a circle contain the abnormal area. The same procedure is performed on normal mammogram images with a random selection location to extract normal ROI. Fig. 2 illustrates the ROI extraction process. In this work, all extracted ROI are resized into 128x128 pixels for calculation simplicity and uniformity of the ROIs. Thus in this phase, the rectangular ROIs are extracted and the ROIs are free from the unwanted area, artifacts and noise which is defined by equation (1) [11] and Fig. 3 shows the extracted ROIs, which containing different abnormalities.

$$I_{ROI} = I[X_c - r, (1024 - Y_C) - R, 2\times r, 2\times r] \quad (1)$$

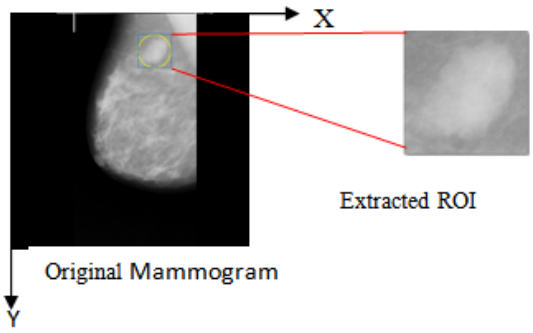


Fig. 2.ROI extraction process from mammographic digital image.

B. Gray-Level Co-occurrence Matrix (GLCM)

Grey level co-occurrence matrices (GLCM) are introduced in [12, 13]. It explains the occurrence of certain grey levels in relation to other grey levels using statistical sampling. The process is explained in the following paragraph.

Assume that an image to be analyzed is rectangular and has N_x rows and N_y columns. The gray level appearing at each

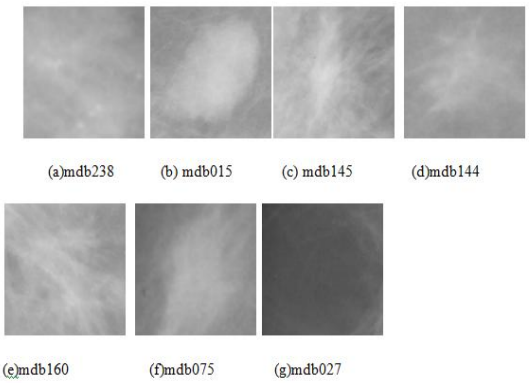


Fig. 3.Extracted ROIs of MIAS database: (a) Calcification (b) Well-defined mass (c) Spiculated mass (d) Ill-defined mass (e) Architectural distortion (f) Asymmetrical mass (g) normal

pixel is quantized to N_g levels. Let, $L_x = \{1, 2, ...N_x\}$ be the rows, $L_y = \{1, 2, ...N_y\}$ be the columns and G={0,1,2,....N_g-1} is the total number of gray levels quantized up to $N_g$levels. The set $L_x \times L_y$ is the set of pixels of the image ordered by their row-column designations. Then, the image I can be represented as a function of co-occurrence matrix that assigns some gray level in $L_x \times L_y$ as $I : L_x \times L_y \rightarrow G$.

The texture-context information is specified by the matrix of relative frequencies P_{ij} with two neighbouring pixels separated by distance d, one with gray level i and the other with gray level j. Such matrices of gray-level co-occurrence frequencies are a function of the angular relationship θ and distance d between the neighbouring pixels. By using a distance of one pixel and angles quantized to 45^0 intervals, four matrices of horizontal, first diagonal, vertical, and second diagonal (0, 45, 90 and 135 degrees) are used. Then, the un-normalized frequency in those four directions is defined by equation (2).

$$p(i, j, d, \theta) = \#$$
$$\begin{cases} ((k,l),(m,n)) \in |(L_x \times L_y) \times (L_x \times L_y)| \\ (k-m=0, |l-n| = d) or (k-m=d, l-n=-d) \\ or(k-m=-d, l-n=d) or(|k-m| = d, l-n=0), \\ or(k-m=d, l-n=d) or(k-m=-d, l-n=-d), \\ I(k,l)=i, \quad I(m,n)=j \end{cases} \quad (2)$$

Where # is the number of elements in the set, *(k, l)* the coordinates with gray level *i*, *(m, n)* the coordinates with gray level *j*.

C. Feature extraction using curvelet coefficient

In this work, texture features are extracted to represent digital mammographic ROI images. The digital mammographic ROIs possess various line and curve singularities, which may contain significant information. We need to explore these geometric properties of the complex structure i.e, line and curve information of ROI images. Using multi-resolution approach, we can analyze the complex structure of an image at various level.

Wavelet transform is one of the most widely accepted multi-resolution approach used by different researchers due to its properties of time frequency localization and multi-resolution. Due to its lower directionality and isotropic scaling, it represents well the 1D point singularities but, it overlooks the smoothness of the 2D singularities of complex structure [14]. To overcome the drawbacks of the conventional wavelet transform, a new multi-resolution geometric analysis tool was introduced in [15] named curvelet transform. The curvelet transform is derived from ridgelet transform, which can represent only line singularities, due to that it was not popular with many image and signal processing problems [16]. But, when the ridgelet transform is applied to the sub-images, it represents the curve singularities in more effectively. In this process, the original image decomposes into subbands followed by the spatial partitioning of each subblock, which is known as the first-generation curvelet transform. It gains less popularity due to its indistinct geometry of ridgelets. To address the limitations of the first generation curvelet, Candes et al. [17] proposed a second-generation curvelet, namely fast discrete curvelet transform (FDCT) using fast fourier transform for computational simplicity. There exists two implementation procedure of FDCT, the first one is based on unequally spaced fast fourier transforms (USFFT), whereas the second one is based on the wrapping function. We chose wrapping based FDCT, because of its simplicity, faster implementation and also it decreases the repetition of information [17, 18].

The discrete curvelet coefficient of a 2D image is computed as a cartesian array $f[m,n]$, with $0 \le m < M, 0 \le n < N$, which is defined by the equation (3).

$$c^D(j,l,k_1,k_2) = \sum_{m=1}^{M}\sum_{N=1}^{N} f[m,n]\varphi^D_{jlk_1k_2}[m,n] \qquad (3)$$

Where, $\varphi_{jlk_1k_2}$ is the digital curvelet waveform and j, l, (k_1,k_2) are the output parameters of scale, orientation and locations.

In this work, we applied a wrapping based FDCT method on mammographic ROI images to extract texture features. In this experiment, each 128 x 128 ROI image is decomposed into multiple subband as a curvelet coefficients. Figure 4 shows the output of curvelet coefficients for ROI image at 5 scales. Every scale holds texture information in different directions except the first and last scale. After 5 scale decomposition, 82 = (1+16+32+32+1) subbands are generated as curvelet coefficients, which are used for feature extraction. Among them, scale 1 represents the approximation component and others are detail components. In the present work, two types of feature vectors have been generated, first one from the approximation coefficients and the second one is directional energy feature which has been calculated using equation (4) from each subband to form another feature vector. Consider p(i,j) be the (i.j)th entry in a normalized GLCM. G is the number of gray levels range from 0 to N_{g-1}

$$Energy = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} \{p(i,j)\}^2 \qquad (4)$$

The size of the approximation subband is found to be 11x11 pixels, thereafter pixels values are converted as row major order to produced feature vector and therefore the length of the feature vector is 121 for approximation subband. The length of the energy features vector is 82 for all subband coefficients. The length of the feature needs to be reduced to find suitable and effective features.

D. Feature selection

Feature selection is very much important to obtain greater accuracy. Too many features many not give greater classification accuracy. Feature optimization is required to get better results. After feature extraction, statistical based two-sample t-test and F-test have been utilized to select more effective features from the feature matrix, which give good results in our classification problem. In this method, the values of t and F can be calculated by the equations (5) and (6).

$$t = \frac{\left|\mu_{v_1} - \mu_{v_2}\right|}{\sqrt{\frac{(\delta_{v_1})^2}{N_{v_1}} + \frac{(\delta_{v_2})^2}{N_{v_2}}}} \qquad (5)$$

$$F = \frac{s^2_{v_1}}{s^2_{v_1}} \qquad (6)$$

Where, N_{v1} and N_{v2} are the numbers of ROIs in two classes. Here, μ_{v1} and μ_{v2} are mean, σ_{v1} and σ_{v2} are standard deviations, and S_{v1} and S_{v2} are the variances of two classes. A higher *t* and *F* value indicates more significant differences between the means of the two vectors.

III. Experimental Results and Discussion

The experiments have been conducted on Matlab and tested on MIAS mammographic database images [10]. To evaluate the effectiveness of the proposed method, we compute various performance metrics such as sensitivity, specificity, accuracy and ROC. These metrics can be derived from a confusion matrix shown in Table I. A confusion matrix is a descriptive term that represents the performance of the classifier in a tabular layout of the actual and the predicted class. Sensitivity, specificity and accuracy can be defined by the equations (7), (8) and (9).

. Table- I: Confusion matrix for two-class problem with different performance measures

Output Class	Confusion Matrix	
Positive	*True Positive (TP)*	*False Positive (FP)*
Negative	False Negatoive (FN)	True Negative (TN)

$$Sensitivity = \frac{TP}{TP + FN} \qquad (7)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (8)$$

$$Accuracy = \frac{TP + TN}{Total\ \ Number\ \ of\ \ Sample} \qquad (9)$$

.

Table- II: Different number of mammographic images used during classification.

		Number of images used		
Image class	Total no. of images	Training	Testing	Validation
Normal-Abnormal	322	226	48	**48**
Benign–malignant	115	81	17	17

Table III- Performance measures (%) of the classifier for normal and abnormal class using approximation subband feature and T-test at different significance level (α)

Significance Level (α)	Feature Selected	Sensitivity	Specificity	Accuracy
0.9	114	92.3	78.3	**87.3**
0.7	102	93.8	62.5	83.3
0.5	88	89.7	66.7	85.4
0.3	74	96.4	70.0	85.4

Table IV- Performance measures (%) of the classifier for Benign-Malignant class using approximation subband feature and T-test at different significance level (α)

Significance level (α)	Feature Selected	Sensitivity	Specificity	Accuracy
0.9	113	72.7	50.0	64.7
0.7	91	54.5	83.3	64.7
0.5	64	77.8	25.0	52.9
0.3	30	88.9	75.0	**82.4**

Table V- Performance measures (%) of the classifier for normal and abnormal class using approximation subband feature and F-test at different significance level (α)

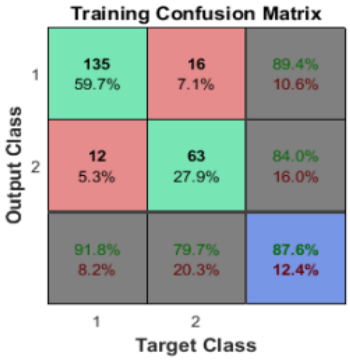
Significance level (α)	Feature Selected	Sensitivity	Specificity	Accuracy
0.9	118	88.7	79.1	**86.9**
0.7	116	90.0	83.3	87.5
0.5	113	96.7	66.7	85.4
.0.3	107	89.3	60.0	77.1

Table VI- Performance measures (%) of the classifier for Benign-Malignant class using approximation subband feature and F-test at different significance level (α)

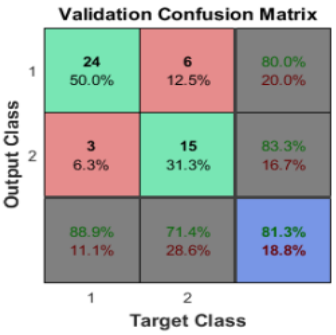
Significance level (α)	Feature Selectd	Sensitivity	Specificiy	Accuracy
0.9	112	71.9	60.8	**67.0**
0.7	91	75.0	33.3	52.9
0.5	65	66.7	80.0	70.6
0.3	37	77.8	37.5	58.8

Table VII- Performance measures (%) of the classifier using energy feature of all curvelet coefficients and t-test method

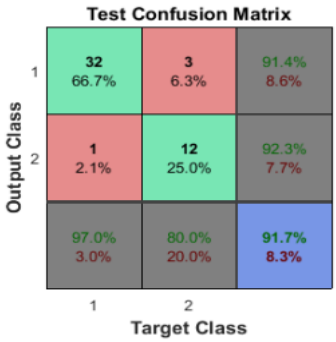
Class	Sensitiviy	Specificiy	Accuray
Normal-Abnormal	96.13	93.91	**95.34**
Benign- Malignant	85.93	74.50	**80.86**



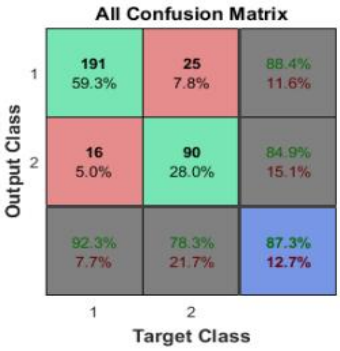
(a)



(b)



(c)



(d)

Fig.4. (a-d) Confusion matrices of classifier for classification of mammograms as normal and abnormal using approximation subband feature.

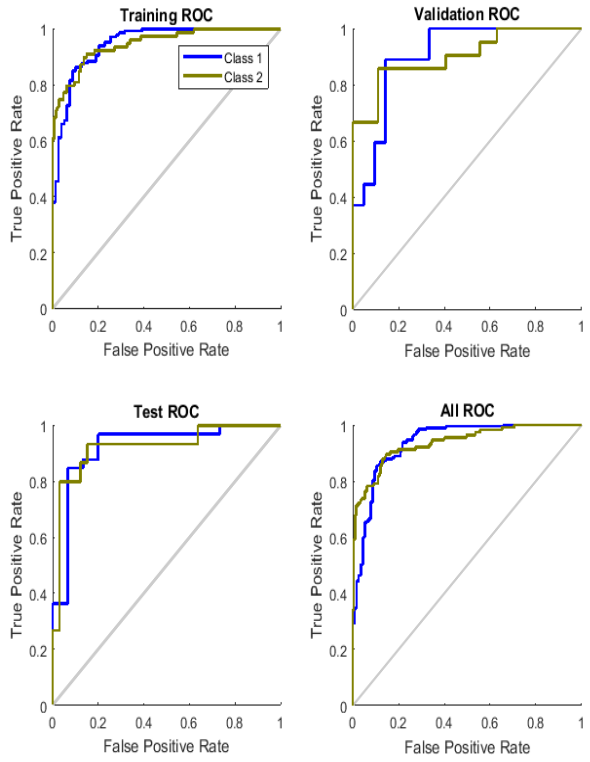


Fig. 5. Comparison of ROC curves of mammogram classification as normal and abnormal using curvelet approximation subband coefficient and t-test method.

IV. CONCLUSION

The CAD system is developed and presented here for the classification of mammogram into normal and abnormal breast tissues with the aim to support the radiologists in visual diagnosis. The proposed method utilizes curvelet transform and calculated approximation subband coefficient feature and energy feature form mammograms ROI image. From the experimental results it is observed that, t-test based energy features achieves higher classification accuracy with ANN classifier than that of F-test. The experimental results show that, an accuracy of 95.34% and 80.86% have been obtained for normal-abnormal and benign-malignant respectively in MIAS database.

REFERENCES

1. Globocan project 2012, International Agency for Research on Cancer (iarc), World Health Organisation: cancer fact sheets, http://globocan.iarc.fr.
2. L. Tabar, P. Dean, Mammography and breast cancer: the new era, International Journal of Gynecology & Obstetrics, 82 (3) pp. 319–326, 2003.
3. H. Cheng, X. Shi, R. Min, L. Hu, X. Cai, H. Du, Approaches for automated detection and classification of masses in mammograms, Pattern recognition, 39 (4), pp. 646–668, 2006.
4. S. Beura, B. Majhi, and R. Dash, Mammogram Classification using Two Dimensional Discrete Wavelet Transform and Gray-Level Co-occurrence Matrix for Detection of Breast Cancer, Neurocomputing, vol. 154, pp. 1-14, 2015.
5. A. Dhawan et al, Analysis of mammographic microcalcifications using grey-level image structure features, IEEE Transactions on Medical Imaging 15 (3), 246–259, 1996.
6. D. Wei, H.-P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, M. M. Goodsitt, False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis, Medical Physics 24 (6), 903–914, 1997.
7. S. Liu, C. F. Babbs, E. J. Delp, Multiresolution detection of spiculated lesions in digital mammograms, Image Processing, IEEE Transactions on 10 (6), 874–884, 2001.
8. E. A. Rashed, I. A. Ismail, S. I. Zaki, Multiresolution mammogram analysis in multilevel decomposition, Pattern Recognition Letters 28 (2), pp. 286–292, 2007.
9. P. Gorgel, A. Sertbas, O. N. Ucan, Mammographical mass detection and classification using local seed region growing spherical wavelet transform (lsrg–swt) hybrid scheme, Computers in biology and medicine 43 (6), 765–774, 2013.
10. P. SUCKLING J, "The mammographic image analysis society digital mammogram database," Digital Mammo, pp. 375–386, 1994.
11. R. Biswas, A. Nath, S. Roy, Mammogram classification using gray-level co-occurrence matrix for diagnosis of breast cancer, 2016 International Conference on Micro-Electronics and Telecommunication Engineering, pp. 161-166, 2016. DOI 10.1109/ICMETE.2016.85
12. Robert M. Haralick, K. Shanmugam, and ITS'HAK Dinstein, *"Textural features for image classification",* IEEE Transaction on system, man and cybernetics, vol.SMC-3, pp. 610-621,No.6, November 1973
13. L.K. Soh, and C. Tsatsoulis, "Texture analysis of sar sea ice imagery using grey level co-occurrence matrices", IEEE Transactions on Geoscience and Remote Sensing.
14. D. R. Nayak, R. Dash, B. Majhi, and V. Prasad, "Automated pathological brain detection system: A fast discrete curvelet transform and probabilistic neural network based approach," Expert Systems with Applications, vol. 88, pp. 152–164, 2017.
15. E. J. Candes and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," tech. rep., Stanford Univ Ca Dept of Statistics, 2000.
16. M. N. Do and M. Vetterli, "The finite ridgelet transform for image representation," IEEE Transactions on image Processing, vol. 12, no. 1, pp. 16–28, 2003.
17. E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," Multiscale Modeling & Simulation, vol. 5, no. 3, pp. 861–899, 2006.
18. N. Gedik and A. Atasoy, "A computer-aided diagnosis system for breast cancer detection by using a curvelet transform," Turkish Journal of Electrical Engineering & Computer Sciences, vol. 21, no. 4, pp. 1002–1014, 2013.

AUTHORS PROFILE

Ranjit Biswas received his B.Tech. degree in Information Technology in 2007 and M.Tech. degree in Information Technology in 2009. Currently, he is pursuing Ph.D degree at Assam University, Silchar. Currently, he is working as an Assistant Professor in the Department of Information Technology at Ramkrishna Mahavidyalaya, Kailashahar, Unakoti Tripura, India. His areas of interest are digital image processing and content based image retrieval.

Sudipta Roy is currently serving as Professor in the department of CSE, TSSOT, Assam University. He is presently holding the post of Dean, TSSOT. He has received his MCA from BITS, Mesra and ME degree from Jadavpur University, Kolkata. He received his PhD from Assam University, Silchar, in the year 2010. He has guided many PhD theses and also has completed a number of research projects and Industrial projects. He also have a huge number of national and international publication. His research interest includes, image processing, soft computing, AI and neural networks to name a few.

ABHIJIT BISWAS is currently working as an Assistant Professor in the department of CSE, TSSOT, Assam University since October 2009. He is a alumni of NIT Bhopal and have received his M.Tech and PhD degrees from NERIST in the year 2009 and 2018 respectively. His research interest includes, network routing, Network – On – Chip topologies and Data Science.