Texture Based Segmentation using Statistical Properties for Mammographic Images

H.B.Kekre

Department of Computer Engineering
Senior Professor, MPSTME, SVKM's NMIMS University
Mumbai, India
hbkekre@yahoo.com

Saylee Gharge

Department of Electronics and telecom. Engineering
Ph.D. Scholar, MPSTME, SVKM's NMIMS University
Mumbai, India
Sayleegharge73@yahoo.co.in

Abstract— **Segmentation is very basic and important step in computer vision and image processing. For medical images specifically accuracy is much more important than the computational complexity and thus time required by process. But as volume of data of patients goes on increasing then it becomes necessary to think about the processing time along with accuracy. Here in this paper, new algorithm is proposed for texture based segmentation using statistical properties. For that probability of each intensity value of image is calculated directly and image is formed by replacing intensity by its probability . Variance is calculated in three different ways to extract the texture features of the mammographic images. These results of proposed algorithm are compared with well known GLCM and Watershed algorithm.**

Keywords- Segmentation, Variance, Probability, Texure

I. INTRODUCTION

Segmenting mammographic images into homogeneous texture regions representing disparate tissue types is often a useful preprocessing step in the computer-assisted detection of breast cancer. With the increasing size and number of medical images, the use of computers in facilitating their processing and analysis has become necessary. Estimation of the volume of the whole organ, parts of the organ and/or objects within an organ i.e. tumors is clinically important in the analysis of medical image. The relative change in size, shape and the spatial relationships between anatomical structures obtained from intensity distributions provide important information in clinical diagnosis for monitoring disease progression. Therefore, radiologists are particularly interested to observe the size, shape and texture of the organs and/or parts of the organ. For this, organ and tissue morphometry performed in every radiological imaging centre. Texture based image segmentation is area of intense research activity in the past few years and many algorithms were published in consequence of all this effort, starting from simple thresholding method up to the most sophisticated random field type method. The repeating occurrence of homogeneous regions of images is texture. Texture image segmentation identifies image regions that have homogeneous with respect to a selected texture measure. Recent approaches to texture based segmentation are based on linear transforms and multiresolution feature extraction [1], Markov random filed models [2,3], Wavelets [4 – 6] and fractal dimension [7]. Although unsupervised texture-based image segmentation is

not a novel approach, these have limited adoption due to their high computational complexity.

Segmentation methods are based on some pixel or region similarity measure in relation to their local neighborhood. These similarity measures in texture segmentation methods use some textural spatial-spectral-temporal features such as Markov random field statistics (MRF) [8-10], co-occurrence matrix based features [11], Gabor features [12], local binary pattern (LBP) [13], autocorrelation features and many others. A number of image processing methods have been proposed to perform this task. S. M. Lai et al. [14] and W. Qian et al. [15] have proposed using modified and weighted median filtering, respectively, to enhance the digitized image prior to object identification. D. Brzakovic et al. [16] used thresholding and fuzzy pyramid linking for mass localization and classification. Other investigators have proposed using the asymmetry between the right and left breast images to determine possible mass locations. Yin *et al.* [17] uses both linear and nonlinear bilateral subtractions while the method by Lau *et al.* [18]. relies on "structural asymmetry" between the two breast images. Recently Kegelmeyer has reported promising results for detecting spiculated lesions based on local edge characteristics and Laws texture features [19-21].The above methods produced a true positive detection rate of approximately 90%. Various segmentation techniques have been proposed based on statistically measurable features in the image [22-27] Clustering algorithms, such as k-means and ISODATA, operate in an unsupervised mode and have been applied to a wide range of classification problems.

For this paper gray level co-occurrence matrix based features and watershed algorithm are considered for comparison with proposed algorithm which is based on statistical properties for segmentation of mammographic images. In section II different algorithms for texture based segmentation are explained in detail. Section III shows results for those methods and section IV concludes the work.

II. ALGORITHMS FOR TEXTURE BASED SEGMENTATION

Texture is one of the most important defining characteristics of an image. It is characterized by the spatial distribution of gray levels in a neighborhood. In order to capture the spatial dependence of gray-level values which contribute to the perception of texture, a two dimensional dependence texture analysis matrix are discussed for texture

consideration. Since texture shows its characteristics by both each pixel and pixel values. There are many approaches using for texture classification.

A. Gray Level Co-occurrence Matrix(GLCM)

The gray-level co-occurrence matrix seems to be a well-know statistical technique for feature extraction. However, there is a different statistical technique using the absolute differences between pairs of gray levels in an image segment that is the classification measures from the Fourier spectrum of image segments. Haralick suggested the use of gray level co-occurrence matrices (GLCM) for definition of textural features. The values of the co-occurrence matrix elements present relative frequencies with which two neighboring pixels separated by distance *d* appear on the image, where one of them has gray level *i* and other *j*. Such matrix is symmetric and also a function of the angular relationship between two neighboring pixels. The co-occurrences matrix can be calculated on the whole image, but by calculating it in a small window which scanning the image, the co-occurrence matrix can be associated with each pixel.

For a 256 gray levels image one should compute 256x256 co-occurrence matrices at all positions of the image. It is obvious that such matrices are too large and their computation becomes memory intensive. Therefore, it is justified to use a less number of gray levels, typically 64 or 32. There is no unique way to choose the values of distance, angle and window, because they are in relationship with a size of pattern.

Using co-occurrence matrix textural features are defined as:

Maximum Probability: $\max(P_{ij})$

$$\text{(1.1)}$$

$$\text{Variance} = (\sum(i-\mu_i)^2 \sum P_{ij})(\sum(j-\mu_j)^2 \sum P_{ij}) \tag{1.2}$$

$$\text{Correlation} = \sum_i \sum_j (i-\mu_x)(j-\mu_y)P_{ij}/\sigma_x\sigma_y \tag{1.3}$$

where μ_x and μ_y are means and σ_x, σ_y are standard deviation

$$\text{Entropy} = \sum_i \sum_j P_{ij}\log(P_{ij}) \tag{1.4}$$

$$\text{Energy} = \sum_i \sum_j p_{i,j}^2 \tag{1.5}$$

Amongst all these features variance, probability has given the best results. Hence results for these extracted features using gray level co-occurrence matrix are displayed in section III.

B. Watershed Algorithm

The watershed transformation is a powerful tool for image transformation .Beucher and Lantuejoul were the first to apply the concept of watershed and divide lines to segmentation problems [28].They used it to segment images of bubbles and metallographic pictures. The watershed transformation considers the gradient magnitude of an image as a topographic

surface. Watershed segmentation [29] classifies pixels into regions using gradient descent on image features and analysis of weak points along region boundaries. The image feature space is treated, using a suitable mapping, as a topological surface where higher values indicate the presence of boundaries in the original image data. It uses analogy with water gradually filling low lying landscape basins. The size of the basins grow with increasing amounts of water until they spill into one another. Small basins (regions) gradually merge together into larger basins. Regions are formed by using local geometric structure to associate the image domain features with local extremes measurement.

Watershed techniques produce a hierarchy of segmentations, thus the resulting segmentation has to be selected using either some prior knowledge or manually with trial and error. Hence by using this method the image segmentation can not be performed accurately and adequately, if we do not construct the objects we want to detect. These methods are well suited for different measurements fusion and they are less sensitive to user defined thresholds. In this approach, the picture segmentation is not the primary step of image understanding. On the contrary, a fair segmentation can be obtained only if we know exactly what we are looking for in the image. For this paper, watershed algorithm for mammographic images is implemented as mentioned in [30] and displayed as a result in Figure1(b) and 2(b) .

C. Proposed Algorithm

From the previous section it can be inferred that even though variance using GLCM gives proper tumor demarcation for mammographic images it require huge computation time to calculated statistical properties for the image. Watershed algorithm is comparatively less complex hence less computation time is required but this method gives over segmentation. Hence to achieve proper segmentation with less complexity, new algorithm has been proposed. In this proposed algorithm statistical properties such as variance, probability for grouping pixels into regions and then images are formed for each statistical property.

1) Probability

Images are modeled as a random variable. A full understanding of the properties of images and of the conclusions has drawn from them thus demand accurate statistical models of images. In this paper, probability of image is considered for extraction of the features of an image.

For complete image, probability of particular i^{th} gray level is calculated which is given by:

$$\text{Probability } P(i) = \frac{X_i}{MXN} \tag{1.6}$$

Where Xi is number of pixels for i th gray levels, M and N are no. of rows and columns of the image.

After calculating this the image is formed which contains probability values for that particular gray level instead of gray level in the image .Since the values of probabilities are too

small they are invisible. For perceptibility of this image histogram equalization is preferred and displayed as equalized probability image as shown in Figure 1(e) and 2(e) for mammographic images.

2) Variance

Variance is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean. The variance of a real-valued random variable is its second central moment, and it also happens to be its second cumulant. The variance of random variable is the square of its standard deviation.

$$\text{Variance } (X) = E\ [\ (\ X - \mu)^2\ ] \qquad (1.7)$$

if $\mu = E(X)$, where $E(X)$ is the expectation (mean) of the random variable X. That is, it is the expectation of the square of the deviation of X from its own mean. It can be expressed as "The average of the square of the distance of each data point from the mean", thus it is the mean squared deviation. This same definition is followed here for images. Initially probability of image is calculated. Since the probability values are very small, equalized probability image is applied as an input image to find variance of probability image by using 3x3 window size as given by Equation 1.7. Results are shown in the section III.

By using this approach any abnormality in the image can be observed very easily but quite often the radiologist need

other details. In this case original image is used instead of using equalized probability image as an input image. Thus variance of original image is calculated using same Equation 1.7 for window size 3x3 and results are shown as direct variance image in the section III. In the third approach , variance is calculated using probability of the image as given by equation 1.8.Results are shown as variance using probability Figure1(h) and 2(h)

$$\text{Variance using probability } (X) = E\ [\ (\ X - \mu)^2\ x\ P(X)\ ] \quad (1.8)$$

III. RESULTS

Mammography images from mini-mias database were used in this paper for implementation of GLCM, Watershed and proposed algorithm for tumor demarcation. Fig.2 (a) shows original image with tumor. It has fatty tissues as background. Class of abnormality present is CIRC which means well-defined/ circumscribed masses. Image 1 and Image 2 (mdb184 and mdb028 from database) have malignant abnormalities.Figure 1(a) and 2 (a) show original mammographic images. Figure 1(b) and 2 (b) indicates segmentation using watershed algorithm .Figure 1 and 2 (c)-(d) show results for probability and variance using GLCM. Figure 1 and 2 (e)-(h) show equalized probability, variance of probability ,direct variance and variance using probability image for image 1 and image 2.
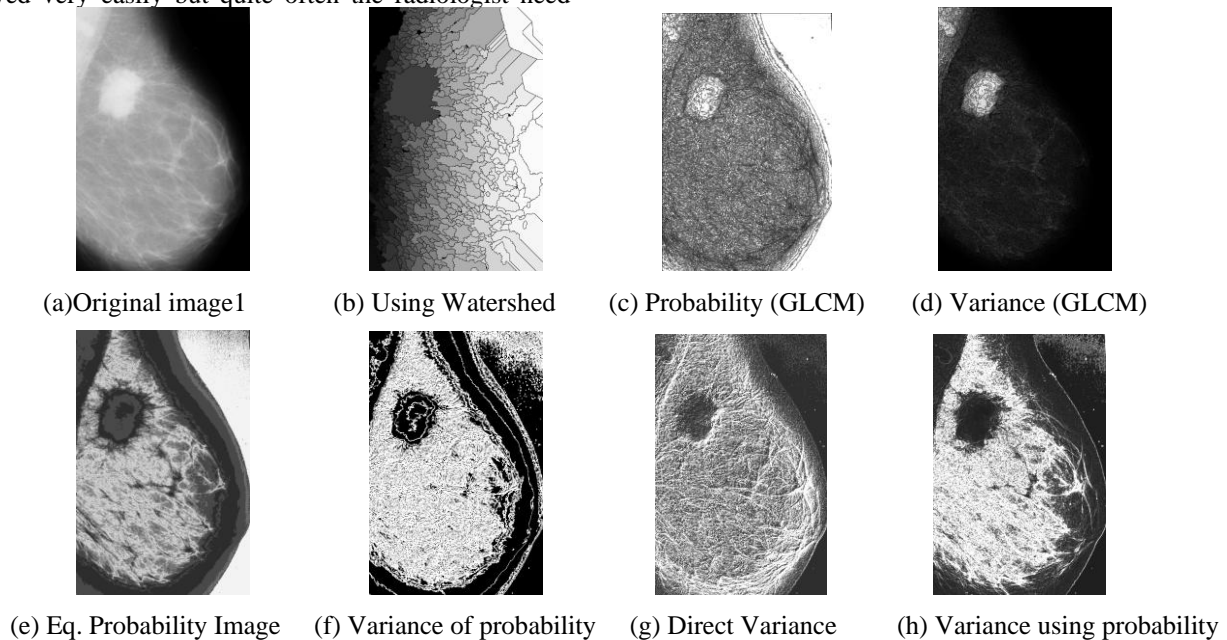

(a)Original image1   (b) Using Watershed   (c) Probability (GLCM)   (d) Variance (GLCM)

(e) Eq. Probability Image   (f) Variance of probability   (g) Direct Variance   (h) Variance using probability

Figure 1 Results of Watershed, GLCM and proposed algorithm for image 1

(a) Original image2   (b) Using Watershed   (c) Probability (GLCM)   (d) Variance (GLCM)

(e)Eq. Probability Image   (f) Variance of probability   (g) Direct Variance   (h) Variance using probability
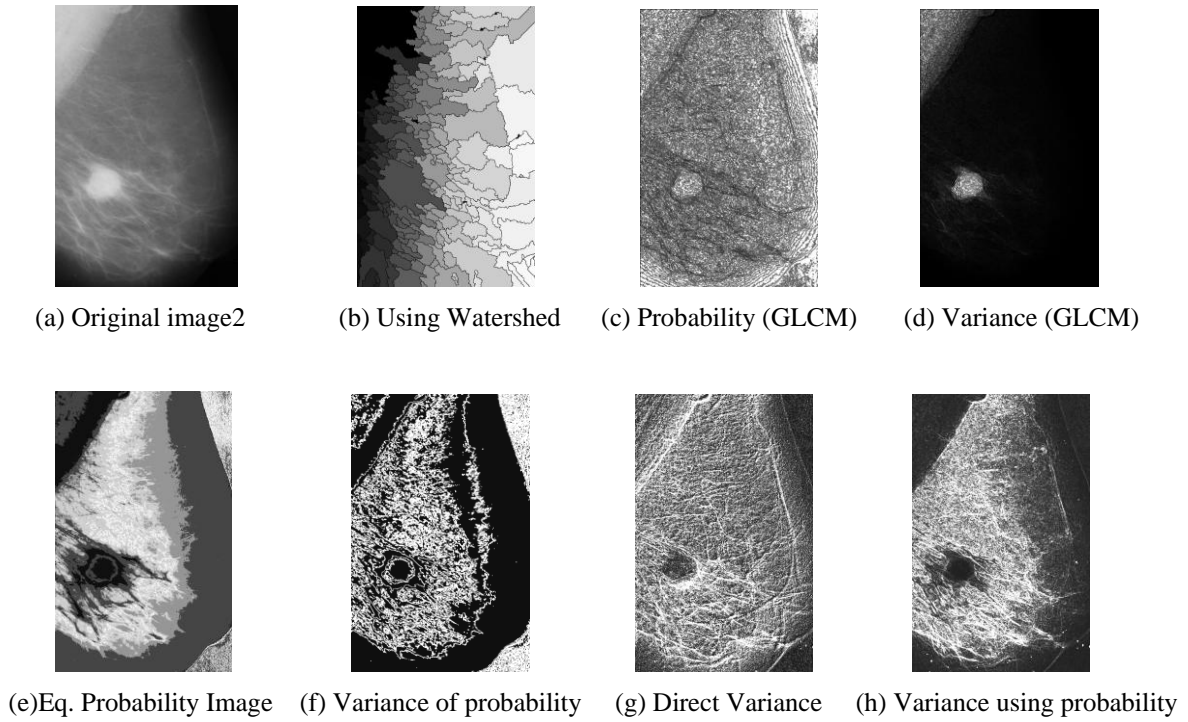
Figure 2.Results of Watershed, GLCM and proposed algorithm for image 2

Table 1: Performance comparison of GLCM, Watershed and Proposed Algorithm

Method	Tumor demarcation	Tumor details	Other details	Remarks
(1)GLCM				
Variance	Clear	Clear	Obscure	Acceptable
Probability	Clear	Clear	Obscure	Acceptable
(2)Watershed Algorithm	Clear	Not clear	Obscure	Over segmentation
(3)Proposed Algorithm				
Variance	Clear	Clear	Clear	**Recommended**
Eq. Probability	Clear	Clear	Obscure	Acceptable

IV.   CONCLUSION

From Table 1 it can be inferred that GLCM method results are not very good but acceptable but have high computational complexity. As far as watershed algorithm is concerned the results are not acceptable because of over segmentation. The results of proposed methods using statistical parameters such as variance, probability are all

acceptable, amongst which direct variance method gives the best results for mammographic images which are verified by radiologist.

REFERENCES

[1] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE transaction on Image Processing*, vol. 4, no. 11, pp. 1549–1560, 1995.

[2] B.Wang and L. Zhang, "Supervised texture segmentation using wavelet transform," in *Proceedings of the 2003 International Conference on Neural Networks and Signal Processing*, 2003, vol. 2, pp. 1078–1082.

[3] Chang-Tsun Li and Roland Wilson, "Unsupervised texture segmentation using multiresolution hybrid genetic algorithm," in *Proc. IEEE International Conference on Image Processing ICIP03*, 2003, pp. 1033–1036.

[4] T.R. Reed and J. M. H. Du Buf, "A review of recent texture segmentation, feature extraction techniques," in *CVGIP Image Understanding*, 1993, pp. 359–372.

[5] A.K. Jain and K. Karu, "Learning texture discrimination masks," *IEEE transactions of Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 195–205, 1996.

[6] Eldman de Oliveira Nunes and Aura Conci, "Texture segmentation considering multi band, multi resolution and affine invariant roughness.," in *SIBGRAPI*, 2003, pp. 254–261.

[7] L.F. Eiterer, J. Facon, and D. Menoti, "Postal envelope address block location by fractal-based approach," in *17th Brazilian Symposium on Computer Graphics and Image Processing*, D. Coppersmith, Ed., 2004, pp. 90–97.

[8] Haindl, M. & Mikeš, S. (2005). Colour texture segmentation using modelling approach. *Lecture Notes in Computer Science*, (3687), 484–491.

[9] Haindl, M. & Mikeš, S. (2006a). Unsupervised texture segmentation using multispectral modelling approach. In Y. Tang, S.Wang, D. Yeung, H. Yan, & G. Lorette (Eds.), *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, volume II pp. 203–206. Los Alamitos: IEEE Computer Society.

[10] Haindl, M. & Mikeš, S. (2007). Unsupervised texture segmentation using multiple segmenters strategy. In M. Haindl, J. Kittler, & F. Roli (Eds.), MCS 2007, volume 4472 of Lecture Notes in Computer Science pp. 210–219.: Springer.

[11] Robert M. Haralick, Statistical and Structural Approaches to Texture, IEEE Proceedings Of vol. 67, no. 5, May 1979.

[12] Thomas P., William E., "Efficient Gabor Filter Design for Texture segmentation," Pattern Recognition 1996,Pattern Rec.Soc.

[13] Ojala, T. & Pietikainen, M. (1999). Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(477-486).

[14] S. M. Lai, X. Li, and W. F. Bischof, " On techniques for detecting circumscribed masses in mammograms," *IEEE Trans. Med. Zmag.*, vol. 8, no. 4, pp. 377-386, Dec. 1989.

[15] W. Qian, L. P. Clarke, M. Kallergi, and R. A. Clark, " Tree-structured nonlinear filters in digital mammography, *IEEE Trans. Med. Imag.*, vol.13, no. 1, pp. 25-36, Mar. 1994.

[16] D. Brzakovic, X. M. Luo, and P. lBzrakovic, "An approach to automated detection of tumors in mammography," *IEEE Trans. Med. Imag.*, vol. 9, no. 3, pp. 233-241, Sept. 1990.

[17] F. F. Yin, M. L. Giger, K. Dol, C. E. Metz, R. A. Vyborny, and C. J. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images,'' *Med. Phys.*, vol. 18, no. 5, pp. 955-963, Sept. 1991.

[18] T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach,' *Comput. Biomed. Res.*, vol. 24, pp.273-295, 1991.

[19] W. P. Kegelmeyer Jr., J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiol.*, vol. 191, no. 2, pp. 331-337, May 1994.

[20] D. Marr and E. Hildreth, "Theory of edge detection," in Proceeding Royal Society, London., vol. 207, pp. 187-217 , 1980.

[21] J. Lunscher and M. P. Beddoes, "Optimal edge detector design: Parameter selection and noise effects," IEEE Trans. Pattem Anal. Machine Intell., vol. 8, no. 2, pp. 154-176, Mar. 1986.

[22] H. B. Kekre , Saylee Gharge, "Segmentation of MRI Images using Probability and Entropy as Statistical parameters for Texture analysis," Advances in Computational sciences and Technology(ACST),Volume 2,No.2,pp:219-230,2009, http://www.ripublication.com/acst.htm

[23] H. B. Kekre , Saylee Gharge , "Selection of Window Size for Image Segmentation using Texture Features," International Conference on Advanced Computing &Communication Technologies(ICACCT-2008) Asia Pacific Institute of Information Technology SD India, Panipat ,08-09 November,2008.

[24] H. B. Kekre , Saylee Gharge , "Image Segmentation of MRI using Texture Features," International Conference on Managing Next Generation Software Applications ,School of Science and Humanities, Karunya University, Coimbatore, Tamilnadu ,05-06 December,2008.

[25] H. B. Kekre , Saylee Gharge , "Statistical Parameters like Probability and Entropy applied to SAR image segmentation," International Journal of Engineering Research & Industry Applications (IJERIA), Vol.2,No.IV, pp.341-353.

[26] H. B. Kekre , Saylee Gharge , "SAR Image Segmentation using co-occurrence matrix and slope magnitude," ACM International Conference on Advances in Computing, Communication & Control (ICAC3-2009), pp.: 357-362, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg. Available on ACM portal.

[27] H. B. Kekre, Tanuja K. Sarode ,Saylee Gharge, "Detection and Demarcation of Tumor using Vector Quantization in MRI Images" ,International Journal of Engineering Science and Technology(IJEST),Volume 2,No.2,pp:59-66,2009.

[28] Serge Beucher and Christian Lantuejoul, "Use of watersheds in contour detection". In *International workshop on image processing, real-time edge and motion detection* (1979).

[29] Leila Shafarenko and Maria Petrou, " Automatic Watershed Segmentation of Randomly Textured Color Images", *IEEE Transactions on Image Processing*, Vol.6, No.11, pp.1530-1544, 1997.

[30] Basim Alhadidi, Mohammad H. et al, " Mammogram Breast Cancer Edge Detection Using Image Processing Function" Information Technology Journal 6(2):217-221,2007,ISSN-1812-5638.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, He has guided several B.E/B.Tech projects, more than 100 M.

E./M. Tech projects and 17 Ph. D. projects. Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. His areas of interest are Digital Signal processing and Image Processing. He has more than 350 papers in National / International Conferences / Journals to his credit. Recently ten students working under his guidance have received best paper awards. Two research scholars under his guidance have been awarded Ph. D. by NMIMS. Currently he is guiding ten Ph.D. students.

Ms. Saylee M. Gharge has received M.E.(Electronics and telecomm.) degree from Mumbai University in 2007, currently Pursuing Ph.D. from Mukesh Patel School of Technology, Management and Engineering, SVKM'S NMIMS University, Vile-Parle (W), Mumbai. She has more than 9 years of experience in teaching. Currently working as a lecturer in Department of Electronics and Telecommunication in Vivekanand Institute of Technology, Mumbai. Her areas of interest are Image Processing, Signal Processing. She has 32 papers in National /International Conferences/journal to her credit.