

A Systematic Examination of Microarray Segmentation Algorithms

Karthik S. A., Manjunath S. S., Shrinivasa G., Sneha C. R.

Abstract: Microarray is a significant tool and influential method which is used to analyze the cDNA expression in living beings. With the help of this technology one can compute gene expression profile in massive and parallel way. Microarray image segmentation offers an input for subsequent analysis of the extracted microarray data. This work addresses on the different approaches used for segmentation of microarray images. Based on the morphology, topology of spots various methods such as circular shaped, region based, active-contour model based segmentation, shape based, supervised learning and watershed segmentation has been taken for this study. This paper explores and compiles various non statistical approaches used in the field of microarray image segmentation. Finally general tendencies in microarray image segmentation are presented.

Keywords: Microarray, Mean Absolute Error, Spots, Supervised Learning.

I. INTRODUCTION

Classifying an image into forefront area and background area is known as segmentation. In the context of DNA microarray images, segmentation's goal is to divide each grid cell into regions corresponding to the spot (foreground) and background. Segmentation is typically performed after the microarray image has undergone a gridding step and it has been divided into cells (also known as sub-grids).

After segmentation, with the help of foreground area levels of gene expression are estimated. Segmentation is said to be a challenging task due to number of factors. To name a few, the background is typically contaminated by noise. Morphology and dimension of the spot might vary from one spot to another. The intensity of a spot is not necessarily uniform. Because of non homogeneous hybridization mechanism regions of the spot regions might be broken. Additionally, the quality of DNA microarray images might vary. Figure 1 shows a sample sub grid of microarray image. The background noise and variation in the shape of the spots can be clearly seen in the picture. Figure 2 displays some sample spots of different DNA microarray images. As can be seen there is a wide range of variations in terms of shape and quality of the spots. Consequently, segmentation methods normally necessitate manual input of required parameters to obtain promising results. Due to lack of automation biological conclusions are wrongly interpreted. In fact, it has been shown that the selection of the segmentation technique has significant effects on the outcome of an experiment [2].

Revised Manuscript Received on December 05, 2019.

* Correspondence Author

Karthik S. A., Dept. of ISE at Dayananda Sagar Academy of Technology and Management, Bangalore, India. Email: karthiksa1990@gmail.com

Manjunath S. S., Dept. of Computer Science and Engineering ATME, Mysore, Karnataka Country. Email: manjunath.dsatm@gmail.com

Shrinivasa G., Dept. of Computer Science and Engineering ATME, Mysore, Karnataka India.

Sneha C. R., Dept. of Computer Science and Engineering ATME, Mysore, Karnataka India

Whole image investigation involves three significant phases [3, 37, and 38]. 1. Detection of spots (gridding), where locations are assigned to every single spots. 2. Spot partitioning (segmentation), is the task of classifying forefront section and section covered by background and 3. Assessment of intensity (quantification), is estimation of quality spots in each phase.

For the perfect partitioning of image without compromising the quality there is a great demand for robust approach. From the aforesaid stages, segmentation of spots is not an easy task.

DNA microarray image segmentation methods sub divided into several groups on the basis of characteristics they use to divide the spots from background region. Many efficient segmentation algorithms have been proposed over last decades.

A major problem that affects microarray image partitioning is an irregular distribution of intensity values. Sometimes these intensity values appear to reside in the background or it may be in the forefront area of the spot. This leads to misinterpretation of a pixel classification. Hence choosing a perfect segmentation algorithm is a challenging task.

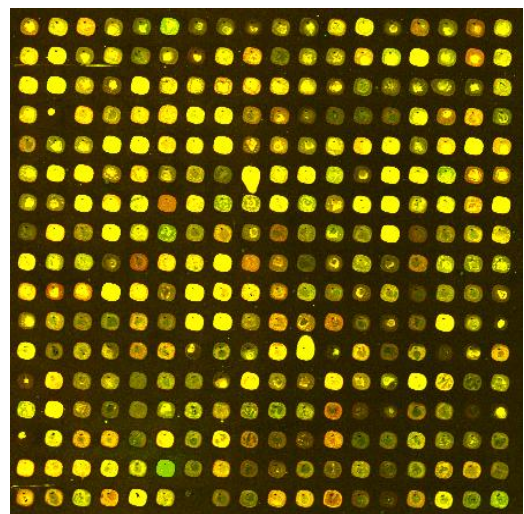


Figure 1: A sample sub grid of microarray image

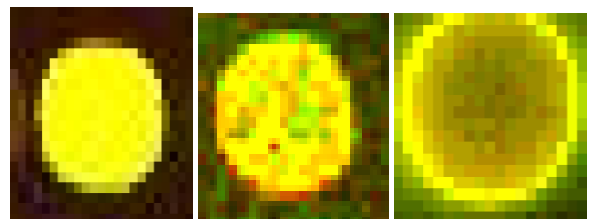


Figure 2: Sample spots of microarray image.

II. MICROARRAY IMAGE SEGMENTATION ALGORITHMS

In this section, different non-statistical DNA microarray segmentation algorithms are reviewed.

A. Circular Shapebased Segmentation

Any prior knowledge regarding the shape of spots can be incorporated into a shape based separation of spots from the background. A simple yet and effective approach for segmentation is fully dependent on prerequisite factors like morphology of the spot must be circular with definite radius. Then a rounded mask of preset radius is displaced within every cell to locate the position of the best match. One of the practical relevance of this algorithm can be observed in a tool called ScanAnalyze [4]. This method has also been incorporated as a choice in numerous software packages such as GenePix [5], ScanArray Express [6], and Dapple [7].

Assuming that each and every one of the spots are spherical with a preset radius is not always a valid hypothesis. An adaptive circle partitioning technique was proposed to address the problems raised by fixed circle segmentation. In this technique each spot is adjusted individually for every cell. Segmentation by adaptive circle can be done in two phases. To begin with, the midpoint of every spot should be assessed. Next, the distance across of the hover must be balanced. This strategy can resolve round shaped spots of various sizes, yet is less viable on unpredictably molded spots. The same concept can be extended to any other shape besides circles.

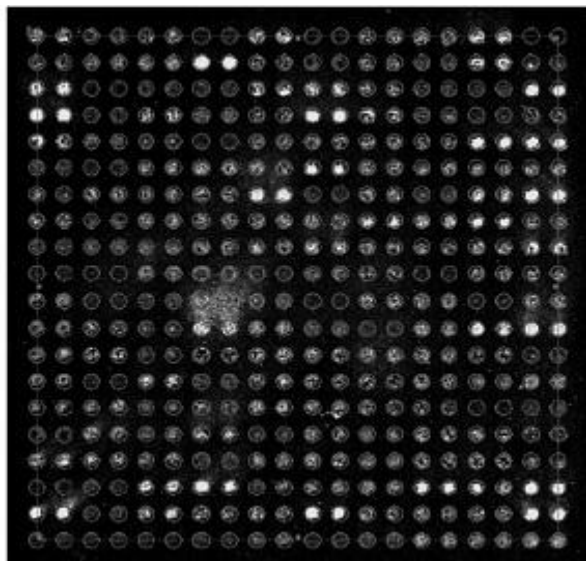


Figure 3: Part of a DNA microarray image segmented using the circular-shape-based approach

B. Region Growing Based Segmentation

Region growing is a classic image segmentation approach that has been used for DNA microarray images [9]. In general, region growing is a process which clusters the pixels into bigger regions based on known criteria for growth [10]. The technique separates every pixel by iteratively divides the region with respect to already known seed points. Often, the seeds consist of a single pixel in the spot (foreground) and a pixel in the background. In each iteration, all the pixels that are not yet labeled but where at least one of their neighboring pixels is labeled are considered. From these pixels, the algorithm adds the most similar pixel (in terms of grey-level)

to the partitioned regions. The algorithm attempts to ensure that the actual fragmented sections are as diverse as possible due to the restriction on connectivity.

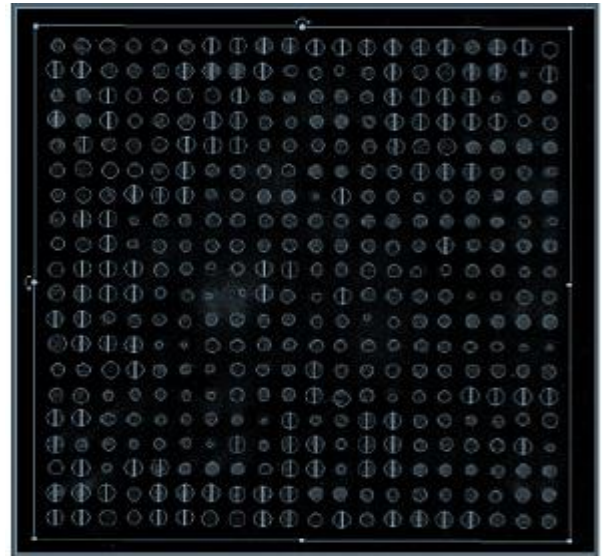


Figure 4: Part of a DNA microarray image segmented using the adaptive circle-based approach

To conclude, the spot's forefront part is decided using area started from the forefront seed and the background area is obtained region from area from the background seed as the background. For microarray segmentation, the foreground and background seeds are chosen based on the outcome of the gridding step. Seeded region growing is employed in a software package named SPOT [9].

C. Clustering Based Segmentation

Classic unsupervised classification algorithms such as the k-means can be employed for microarray image segmentation [13, 14, 15]. Since segmentation is used for separating the spot into two distinct parts and hence number of clusters are also two. The pixels with lowest and highest intensities are chosen as initial clusters. Using a distance measure (e.g., Euclidean distance) every data points are then allocated to the adjacent cluster centers. Thereafter, updated value of cluster center is changed to average intensity of pixel values in every cluster. At last, the algorithm is executed iteratively until the center of the cluster remains unchanged. The approach is interesting since it is self-determining of the contour of the spots. The hybrid k-means microarray segmentation algorithm [16, 33] is a comprehensive edition of the unique k-means segmentation approach. Two most important characteristics of the innovative technique are repetitive clustering and mask harmonizing. If foreground pixels are very few then frequent clustering is applied. In such a situation, the unusual picture elements are detached and process of clustering is continued on picture elements covered by background. The other feature of the hybrid approach is mask matching so as to merge the shape of spot into the algorithm. For every one of the spot all pixels which are in forefront area of the spot of background mask are deleted.

An advantageous outcome of matching of mask is picture elements are partitioned into three clusters rather than two: forefront area, area covered by background, and deleted area. Since it permits a different treatment of artifacts and their removal from further investigation.

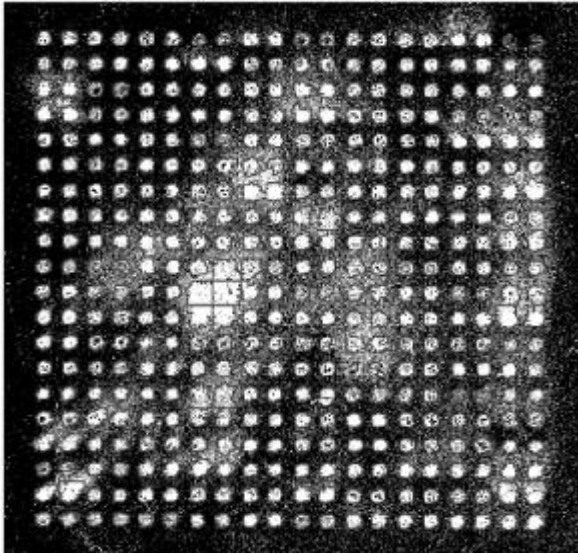


Figure 5: Part of a DNA microarray image segmented using the region growing approach.

D. Criteria Active Contour (Snake) Based Segmentation

The active-contour method used for partitioning of DNA microarray images is based on the active-contour scheme originally proposed [18, 34] is a snake. In this approach the boundaries of the spots are depicted by curves $\Gamma(s)$, that reduces an energy functional consisting of inside components

Generally the snake approach cannot be directly applied to segmentation. The main problem in partitioning is spots might be extremely small and therefore there would be only a small number of pixels along the contour of the spot. This violates the hypothesis of big curve dimension compared to a single picture element.

In boundary-based segmentation, cell area R is assumed to consist of two regions, spot R_1 and background R_2 . It is also assumed that a closed curve separates the cell into $R_1; R_2$ where $R = R_1 \cup R_2$. M_1 and M_2 as the mean value of pixels in R_1 and R_2 , respectively. Let us consider a parametric description of the contour $\Gamma(s)$ where s is the parameter. The total energy encouraged by $\Gamma(s)$ is summation of the snake energies (external) and the spot's energy (internal energy). The total energy is written as

$$E_{\text{total}} = E_{\text{snake}} + \hat{\gamma} E_{\text{region}} \quad (1)$$

$$E_{\text{snake}} = \int_{\Gamma} \left(\frac{\alpha}{2} |\Gamma_s|^2 + \frac{\beta}{2} |\Gamma_{ss}|^2 - \|\nabla I\|^2 \right) ds \quad (2)$$

$$E_{\text{region}} = \frac{\iint_{R_1} (1-M_1)^2 dx dy + \iint_{R_2} (1-M_2)^2 dx dy}{(M_1-M_2)^2} \quad (3)$$

where α and β are constants, $\Gamma_s = \left[\frac{dx}{ds}, \frac{dy}{ds} \right]$

$$\Gamma_{ss} = \left[\frac{d^2x}{ds^2}, \frac{d^2y}{ds^2} \right] A = \pi r^2 \text{ and } \nabla I$$

is the incline of the spotpixel values.

E. Segmentation Using Watersheds

Segmentation by morphological watersheds is a classic image segmentation approach. The concept of watersheds is based on imagination of an image in 3D: two spatial coordinates and intensity [10]. In this perspective, spot is assumed to be an elevation surface in which maximum intensity pixels map to border points and minimum intensity pixels map to gorge points. Assume that this three-dimensional geography is overflowed from beneath and theoretical single-pixel openings are penetrated at every neighborhood minima [10, 26].

F. Supervised Learning based Segmentation

Giannakos et al. [27] introduce a segmentation method that is based on classifying the pixels of a DNA microarray image into foreground (spot), backdrop, and artifact pixels using support vector machines (SVM). Training the data required for the SVM is extracted using the fixed circle technique for real microarray images, whereas for pretended images the training data are extracted during the production of the images. The proposed method consists of the following steps.

- An computerized approach is used to divide every spot of the image into a particular cell. In this cell, vector filters are engaged to remove the noise.
- Next, a pixel-by-pixel classification technique based on SVM is applied.

III. RESULTS AND DISCUSSIONS

Since different DNA microarray segmentation algorithms have been applied to different microarray image sets, it is almost impossible to do a fair comparison of different segmentation algorithms. Moreover, many of the proposed segmentation algorithms have parameters that have to be set properly for achieving the best possible performance. This makes the comparison even more difficult. Additionally, it is very difficult to obtain accurate segmentation results.

The authors of [28, 33, 34, and 35] also point to the fact that result analysis of various segmentation approaches is not a clear-cut task. A group of artificial images were created by a simulator, are used as test images. Different DNA microarray image segmentation methods including fixed circle, adaptive circle, region growing, k-means, and hybrid k-means have been compared which are depicted in table 1. The results indicate that the k-means algorithm gave almost error-free results for the high-quality quality images, For the poor quality DNA microarray images the hybrid k-means algorithm gave the best results. Another approach that the researchers have been using is to compare DNA segmentation algorithms images from replicated experiments. There are two probable methods of replication. First, uses dissimilar microarray substrates for the same experiment. Another choice is to use images where one spot is replicated many times on a single substrate.

Application of different algorithms on images from replicated spot experiments could result in suitable comparison between the segmentation algorithms in terms of consistency. In [22, 32], the authors compare six DNA segmentation methods in terms of mean absolute error (MAE) [30, 31] of replicated spots. Their study shows that the graph-based segmentation method has a smaller MAE compared to circle-shape based and region growing-based segmentation schemes.. The fixed circle method is straightforward to apply, but it fails to segment when the spots are irregular shape and size.. Clustering techniques have the same intrinsic initialization problem.. Watersheds and region growing methods [35] require specification of a primary seed point. In Table 1, a summary of important information for each method, including: 1) segmentation quality reported (might be based on different measures and datasets), 2) the level of automation and the need of parameter adjustment, and 3) associated features and/or limitations.

Table 1: Summary of important segmentation algorithms

Segmentation Method	Segmentation quality	Level of automation	Features/Limitation
Circular	low	High	Limited to circular spots
Region Growing	high	High	Requires foreground and background seed points
Clustering	low	High	Pixels are divided into three groups: foreground, background, and deletions
Active Contour	high	low	Computationally expensive
Watersheds	high	High	Might over-segment the image
Supervised Learning Based	high	High	Requires already segmented images for training

IV. CONCLUSION

With the availability of more powerful computational resources, more advanced techniques will be developed for DNA image segmentation. There is no single method which can be considered good for neither all type of images, nor all methods equally good for a particular type of image. Learning-based segmentation methods should also be considered for DNA image segmentation. However, more research is necessary in the field to assess the efficacy of the method

REFERENCES

- G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
- E. Zacharia and D. Maroulis, "3D spot-modeling for automatic segmentation of cDNA microarray images," *IEEE Transactions on NanoBioscience*, vol. 9, pp. 181–192, November 2010.
- A. A. Ahmed, "Microarray segmentation methods significantly influence data precision," *Nucleic Acids Research*, vol. 32, no. 5, pp. 50–50, 2004.
- M. Katzer, F. Kummert, and G. Sagerer, "Methods for automatic microarray image segmentation," *IEEE Transactions on NanoBioscience*, vol. 2, no. 4, pp. 202–214, 2003.
- M. Eisen, "ScanAlyze, user manual," tech. rep., Stanford University, California, USA, March 1999.
- GenePix Pro 5.0 user's guide and tutorial. Axon Instruments Inc, 2003.
- ScanArray Express Microarray Analysis System User Manual. PerkinElmer Life Sciences, 2002.
- J. Buhler, T. Ideker, and D. Haynor, "Dapple: improved techniques for finding spots on DNA microarrays," Tech. Rep. UWTR 2000-08-05, Department of Computer Science and Engineering, University of Washington, Seattle, WA, August 2000.
- L. Rueda and J. Rojas, "A pattern classification approach to DNA microarray image segmentation," *Pattern Recognition in Bioinformatics*, Lecture Notes in Computer Science, vol. 5780, pp. 319–330, 2009.
- Y. H. Yang, M. Buckley, S. Dudoit, et. al, "Comparison of methods for image analysis on cDNA microarray data," *Journal of Computational and Graphical Statistics*, vol. 11, no. 1, pp. 108–136, 2002.
- R. Gonzalez and R. Woods, *Digital Image Processing*, Third Edition. Prentice Hall, 2008.
- E. R. Dougherty, *Probability and Statistics for the Engineering, Computing, and Physical Sciences*. Prentice-Hall, 1990.
- Y. Chen, E. R. Dougherty, and M. L. Bittner, "Ratio-based decisions and the quantitative analysis of cDNA microarray images," *Journal of Biomedical Optics*, pp. 364–370, 1997.
- D. Bozinov and J. Rahnenfuhrer, "Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering," *Bioinformatics*, vol. 18, pp. 747–756, 2002.
- L. Rueda and L. Qin, "A new method for DNA microarray image segmentation," *Image Analysis and Recognition*, Lecture Notes in Computer Science, vol. 3656, pp. 886–893, 2005.
- G. Weng, Y. Hu, and Z. Li, "cDNA microarray image segmentation using shape-adaptive DCT and k-means clustering," in *2011 International Conference in Electrics, Communication and Automatic Control Proceedings*, pp. 317–324, 2012.
- J. Rahnenfuhrer and D. Bozinov, "Hybrid clustering for microarray image analysis combining intensity and shape features," *BMC Bioinformatics*, vol. 5, no. 47, pp. 1–11, 2004.
- J. Harikiran, D. RamaKrishna, M. Phanendra, et al., "Fuzzy C-means with bi-dimensional empirical mode decomposition for segmentation of microarray image," *International Journal of Computer Science Issues*, vol. 9, pp. 189–198, September 2012.
- M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- J. Ho and W.-L. Hwang, "Automatic microarray spot segmentation using a snake-fisher model," *IEEE Transactions on Medical Imaging*, vol. 27, no. 6, pp. 847–857, 2008.
- J. Gollub, C. A. Ball, G. Binkley, et al., "The Stanford microarray database: Data access and quality assessment tools," *Nucleic Acids Research*, vol. 31, pp. 94–96, Jan. 2003.
- N. Karimi, S. Samavi, S. Shirani, et al., "Segmentation of DNA microarray images using an adaptive graph-based method," *IET Image Processing*, vol. 4, no. 1, pp. 19–27, 2010.
- P. Vesanen, M. Tiainen, and O. Yli-Harja, "Calibration-free methods in segmentation of cDNA microarray images," in *Proc. SPIE Vol. 4667*, p. 291–302, *Image Processing: Algorithms and Systems*, vol. 4667, pp. 291–302, 2002.
- M. Bonner, K. McWeeny, P. Gwynne, et al., "BioChip SNP analysis assay: Development of a 3-D microarray system," *American Journal of Human Genetics*, vol. 67, no. 4, pp. 266, 2000.
- M. Bonner, "BioChip SNP analysis assay: Development of a 3-d microarray system," tech. rep., Motorola BioChip Systems, Arizona, USA.

26. N. Faramarzpour, J. Bondy, and S. Shirani, "Lossless DNA image compression," in Proceedings of the Thirty-Seventh IEEE Asilomar Conference on Signal, Systems and Computers., vol. 2, (Pacific Grove, CA, USA), pp. 1501–1504, 2003.
27. W. Pratt, Digital Image Processing, Third Edition. Wiley, 2001.
28. N. Giannakeas, P. S. Karvelis, T. P. Exarchos, et al., "Segmentation of microarray images using pixel classification—comparison with clustering based methods," Computers in Biology and Medicine, vol. 43, pp. 705–716, July 2013.
29. A. Lehmussola, P. Ruusuvi, and O. Yli-Harja, "Evaluating the performance of microarray segmentation algorithms," Bioinformatics., vol. 22, no. 23, pp. 2910–2917, 2006.
30. Z. Wang, B. Zineddin, J. Liang, et al., "A novel neural network approach to cDNA microarray image segmentation," Computer Methods and Programs in Biomedicine, vol. 11, pp. 189–198, July 2013.
31. StamosKatsigiannis, Eleni Zacharia, and Dimitris Maroulis, "Grow-Cut Based Automatic cDNA Microarray Image Segmentation," *IEEE Transactions On Nanobioscience*, vol. 14(1), 2015.
32. Guifang Shao, Shunxiang Wu, TiejunLi, "cDNAMicroarray Image Segmentation with an Improved Moving K-means Clustering Method," *IEEE ICSC*, 2015.
33. MachmudRAllahamdi, Ito Wasito, "Improved Micro array images Cancer Classificationusing k-nearest neighbour using particle swarm optimization," *IEEE conference, IWBIS*, 2017.
34. S.A. Karthik and S.S. Manjunath " An Automated and Efficient Approach for Spot Identification of Microarray Images Using X-Covariance", Proceedings of International Conference on Cognition and Recognition, Lecture Notes in Networks and Systems 14, Springer Nature Singapore Pte Ltd .pp: 275-282, 2018
35. Karthik S A and Manjunath S S, "An Enhanced Approach for Spot Segmentation of MicroarrayImages," International Conference on Computational Intelligence and Data Science Procedia Computer Science, vol. 132, pp. 226–235, 2018.
36. Yongsheng Pan, Yong Xia, DinggangShen, "Foreground Fisher Vector: Encoding Class-Relevant Foreground to Improve Image Classification," *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, pp. 1-1, 2019.
37. Karthik S A and Manjunath S S , "Automatic Gridding of Noisy Microarray Images Based on Coefficient of Variation", *Informatics in Medicine Unlocked*, vol. 17, 2019
38. Karthik S A and Manjunath S S "Microarray Spot Partitioning ByAutonomously Organizing Maps Through Contour Model", *International Journal of Electrical and Computer Engineering*, vol. 10(1), pp. 746-756, 2020



Dr. Manjunath S. S B.E (CSE), M. Tech(Networking), Ph.D(CSE). He is working as Professor & HOD in Dept. of Computer Science and Engineering ATME, Mysore, Karnataka. He has 17+ years experience in teaching, R&D and Administration. He has published more than 41 papers in International/National Journals and conferences. His areas of interest are image processing and networking



Mr. Shrinivasa G B.E (CSE), M. Tech(CSE). He is working as Asst. Professor in Dept. of Computer Science and Engineering ATME, Mysore, Karnataka. He has 5 years experience in teaching.



Mrs. Sneha C R B.E (ISE), M. Tech(CSE). She is working as Asst. Professor in Dept. of Computer Science and Engineering ATME, Mysore, Karnataka. She has 7.5 years experience in teaching.

AUTHORS PROFILE



Mr. Karthik S A has received B.E (ISE) degree in 2011 and M.Tech (CNE) degree in 2013 from Visvesvaraya Technological University Belgaum, Karnataka, India. Currently he is working as Assistant Professor in Dept. of ISE at Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India and pursuing Ph.D in Visvesvaraya Technological University. He has 6.5 years experience in teaching. His areas of interest include microarray image processing, Data mining and Machine learning.