

Quality Assessment on Satellite Images based on Internal Criterion Techniques

K. M. Sharavana Raju
Dept. of Computer Science
Jazan University
Jazan KSA

Mohammad Shahnawaz Nasir
Dept. of Computer Science
Jazan University
Jazan KSA

T. Meera Devi
Dept. of ECE
Kongu Engg. College
Perundurai, TN, India

ABSTRACT

Segmentation of digital image plays a major role in computer visualization. It is used to extract meaningful objects that exist on the images. Region based clustering is done to extract objects based on the colors present in the satellite images. The principle of clustering is to identify the similar domains from a huge data set to produce an accurate representation of the image. In this paper, k-means, fuzzy c means and kernel fuzzy c means clustering algorithms are used to partition an image data set into number clusters. The images are clustered into four and six categories for which the qualities of the images are compared through the internal criterion techniques Davies–Bouldin index and Dunn index. For this paper, experiment is carried out with more than 100 satellite images. Finally the PASCO Satellite Ortho (PSO) satellite image is selected, which covers the areas around Mt. Kaimondake in Kagoshima, Japan. The experimental results reveal that the quality of the clustered partitions based on the internal criterion conclude, kernel fuzzy c means clustering algorithm performs better than fuzzy c means and k-means clustering methods.

Keywords

Satellite image segmentation, kernel fuzzy clustering, k-means clustering, image quality, Davies–Bouldin index, dunn index internal criterion.

1. INTRODUCTION

Remote sensing is a rapidly developing area in the scientific field and it has a significant extent for the environment. Satellite imagery in remote sensing is based on the resolution of images [3]. The images are of four types - spatial, spectral, temporal and radiometric. The resolution of satellite images varies depending on the sensor used and the altitude of the satellite's orbit. Satellite images have many applications in meteorology, agriculture, geology, forestry, landscape, biodiversity conservation, regional planning, education, seismology, oceanography, intelligence and warfare.

1.1 Segmentation

Image segmentation is the process of image partition into regions with similar attributes [10]. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. In particular, image segmentation is the process of conveying a label to every pixel in an image such that pixels with the same label share certain visual characteristics [1]. The main endeavor of color image segmentation is to identify identical regions in color image that represent objects or meaningful parts of objects present in a scene. Color image segmentation plays an important role in many applications such as environmental assessments such as monitoring of agriculture, monitoring the climate meteorology, object recognition, image compression etc. The

segmentation techniques [5] are classified as pixel-based, region-based, edge-based, fuzzy types and neural networks. In addition to the above hybrid techniques that integrate techniques form different classes. The performance of color segmentation may significantly affect the quality of an image understanding system [4]. The most common features used in image segmentation include texture, shape, grey level intensity and color.

1.2 Clustering

Clustering in image processing is grouping together of pixels from an image, depending on the calculated similarity between them. Clustering can be often defined as an unsupervised classification of pixels [2]. The color image data is naturally clustered in three-dimensional color space (usually RGB). All dominant colors in the image create dense clusters in the color space. Extracting feature based on boundaries on satellite imagery are of great importance in applications like land use land cover (LULC), shore line mapping, boundary delineation and change detection.

Fuzzy c means clustering has the liveness for the pixel that belongs to multiple classes with varying degrees of membership. The clustering approaches can be classified into two general groups: partitional and hierarchical clustering algorithms. Partitional clustering algorithms such as k-means are widely used in many applications such as data mining [11], compression, image segmentation [6], [9] and machine learning. Therefore, the advantage of clustering algorithms is that the classification of objects and easy to implement. Similarly, the drawbacks in this classification are of how to determine the number of clusters and decrease the numbers of iteration [11]. In this paper, k-means clustering is used to find cluster centers that minimize the intra-class variance, i.e., the sum of squared distances from each data point being clustered to its cluster center. Fuzzy c means clustering was used which has the flexibility for the pixel to belong to multiple classes with varying degrees of membership. Kernel Fuzzy c means classify the image based on membership value as well as typicality values.

2. PROPOSED METHODOLOGY

2.1 k-means Clustering

The efficient clustering technique has been developed in the 1960's has been described in classical handbook [2]. In this method a fixed number of clusters and initial cluster centers in the color space are chosen. The key idea is to modify the positions of cluster centers so long as the sum of distances between all points of clusters and their cluster centers will be minimal. During these modifications all points are allocated to closest cluster centers which are the Euclidean distance. After each allocation a new position of cluster centers are computed with the arithmetical means or medians. The iteration stops if the difference between new and old positions is too small. k-

means algorithm is an iterative technique that is used to partition an image into k clusters.

Given a set of observations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, where each observation is a d -dimensional real vector, k -means clustering aims to partition the n observations into k sets ($k \leq n$) $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS)

$$arg_s min \sum_{i=1}^k \sum_{x_j \in S_i} \|X_j - \mu_i\|^2$$

Where, μ_i is the mean of points S_i . The experimental results reveals that k -means clustering has a high Davies–Bouldin index and low Dunn index, which is not suited to assess the quality of an image based on internal criterion, when compared with the other two considered clustering methods viz., the fuzzy c -means and the Kernel Fuzzy c -means.

2.2 Fuzzy C-Means Clustering

Fuzzy C-Means (FCM) is an unsupervised clustering algorithm that has been applied to extensive range of problems involving feature analysis, clustering and classifier design. FCM has a wide domain of applications such as agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis, and target recognition. With the developments of the fuzzy theories, the fuzzy c -means clustering algorithm based on Ruspini fuzzy clustering theory was proposed in 1980s. This algorithm is examined to analyze based on the distance between the various input data points. The clusters are formed according to the distance between data points and cluster centers. The FCM is a method, where the clustering which allows one piece of data to belong to two or more clusters. This method is commonly used in pattern recognition. It is based on minimization of the following objective function:

$$J = \sum_{i=1}^c \sum_{j=1}^N U_{ij}^m \|X_j - c_i\|, 1 \leq m \leq \infty$$

Where,
 m is any real number > 1 ,
 c is number of clusters,
 x_j is the j^{th} object of given N objects,
 U_{ij} is the degree of membership of x_i in the cluster j , c_j is centroid of cluster j ,
 $\|*\|$ is Euclidean distance between any data objects and the centroid.

The parameter m (≥ 1) is called fuzzifier and signifies the amount of fuzziness in the solution set. Fuzzy partitioning is carried out through an iterative optimization as the objective function given in equation, with the update of membership U_{ij} and the cluster centers C_j . The experimental results expose that fuzzy c -means clustering has still a high Davies–Bouldin index and low Dunn index, when compared to the Kernel Fuzzy c -means clustering method, but better than the k -means clustering method which is also not suited to assess the quality of an image based on internal criterion.

2.3 Kernel Fuzzy C-Means Clustering

The 'kernel method' has concerned enormous awareness with the development of support vector machine (SVM). For assessing cluster validity for partitions it is realized by

substitution of a kernel-induced distance metric for the original Euclidean distance, and the corresponding algorithms are called kernel fuzzy c -means (KFCM). By replacing the inner product with an appropriate 'kernel' function, one can implicitly perform a nonlinear mapping to a high dimensional feature space without increasing the number of parameters. This 'kernel method' has been successfully applied into many learning systems, such as Support Vector Machines (SVMs), kernel principal component analysis and kernel fisher discriminant analysis [7].

Given a dataset, $X = \{x_1, \dots, x_n\} \subset R^p$, the original FCM algorithm partitions X into c fuzzy subsets by minimizing the following objective function where c is the number of clusters; n the number of data points; u_{ik} the membership of x_k in class i , satisfying

$$J = \sum_{i=1}^c \sum_{j=1}^N U_{ij}^m \|X_j - c_i\|, 1 \leq m \leq \infty$$

$$\sum_{i=1}^c u_{ik} = 1$$

m the quantity controlling clustering fuzziness, and V the set of cluster centers or prototypes ($v_i \in R^p$). The function J_m is minimized by alternate iterative algorithm. Kernel fuzzy c -means (KFCM) algorithm define a nonlinear map as $\Phi : x \rightarrow \Phi(x) \in F$, where $x \in X$. X denotes the data space and F the transformed feature space with higher even infinite dimension. KFCM minimizes the following objective function

$$J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|\Phi(x_k) - \Phi(v_i)\|^2$$

Where $\|\Phi(x_k) - \Phi(v_i)\|^2 = K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i)$, where $K(x, y) = \Phi(x)^T \Phi(y)$ is an inner product kernel function. If we adopt the Gaussian function as a kernel function, i.e., $K(x, y) = \exp(-\|x - y\|^2 / \sigma^2)$, then $K(x, x) = 1$, can be rewritten as

$$J_m(U, V) = 2 \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m (1 - K(u_k, v_i))$$

The experimental results shows that kernel fuzzy c -means clustering has a very low Davies–Bouldin index and high Dunn index, when compared with the other two considered clustering methods viz., k -means and the fuzzy c -means algorithms, which is the best suited clustering algorithm to assess the quality of an image based on internal criterion.

2.4 Internal Evaluation

Internal evaluation is a method in which the clustering results are evaluated based on the data that was clustered by it. This method is usually assigned the best score to the algorithm that produces clusters with high similarity within a cluster and low similarity between clusters. One disadvantage of using internal criteria in cluster assessment is that high scores on an internal measure do not necessarily result in effective information retrieval applications [8]. Moreover, this evaluation is influenced towards algorithms that use the same cluster model. For example k -Means clustering naturally optimizes object distances and a distance-based internal criterion will likely overrate the resulting clustering.

The following methods are used to assess the quality clustering algorithms based on internal criterion: (a) Davies–Bouldin index (b) Dunn index.

2.4.1 Davies–Bouldin Index

The Davies–Bouldin index can be calculated by the following formula

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right)$$

Where,

n is the number of clusters,

c_x is the centroid of cluster x ,

σ_x is the average distance of all elements in cluster x to centroid c_x ,

$d(c_i, c_j)$ is the distance between centroids c_i and c_j .

Since algorithms that produce clusters with low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity) will have a low Davies–Bouldin index, is considered the best algorithm based on this criterion.

2.4.2 Dunn Index

The Dunn index aims to identify dense and well-separated clusters. It is defined as the ratio between the minimal inter-cluster distance to maximal intra-cluster distance. For each cluster partition, the Dunn index can be calculated by the following formula

$$D = \min_{1 \leq i \leq n} \left\{ \min_{1 \leq j \leq n, i \neq j} \left\{ \frac{d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \right\} \right\}$$

Where,

$d(i, j)$ represents the distance between clusters i and j ,

$d'(k)$ measures the intra-cluster distance of cluster k .

The inter-cluster distance $d(i, j)$ between two clusters may be any number of distance measures, such as the distance between the centroids of the clusters. Similarly, the intra-cluster distance $d'(k)$ may be measured in a variety of ways, such as the maximal distance between any pair of elements in cluster k . Since internal criterion seek clusters with high intra-cluster similarity and low inter-cluster similarity, algorithms that produce clusters with high Dunn index are more desirable.

3. EXPERIMENTAL RESULTS

The program is executed using MATLAB v7.11.0.584 (R2010b). This work presents a novel image segmentation based on color features on the satellite images. First the image is made a color transformation from sRGB to L*a*b. It is then clustered into four and six categories using k-means, fuzzy c means and kernel fuzzy c means clustering algorithms. For this paper, experiment is carried out with more than 100 satellite images. Finally the PASCO Satellite Ortho (PSO) satellite image is selected, which covers the areas around Mt. Kaimondake in Kagoshima, Japan. The outputs are shown in the Figures 1 – 8.

Fig. 1 is the original image, the PASCO Satellite Ortho (PSO) satellite image is selected, which covers the areas around Mt. Kaimondake in Kagoshima, Japan. Fig. 2 shows the image labeled by clustered index in grey scale; Fig. 3 and Fig. 4 shows the four cluster and six cluster outputs of k-means

method; Fig. 5 and Fig. 6 shows the four cluster and six cluster outputs of FCM method; Fig. 7 and Fig. 8 shows the four cluster and six cluster outputs of KFCM method.

The algorithms that produce clusters with low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity) will have a low Davies–Bouldin index, is considered the best algorithm based on this criterion. Since internal criterion seek clusters with high intra-cluster similarity and low inter-cluster similarity, algorithms that produce clusters with high Dunn index are more desirable.

The Davies–Bouldin index and Dunn index values for four clusters and six clusters using the k-means, fuzzy c-means and kernel fuzzy c means clustering algorithms are compared and shown in Table 1 and Table 2.

Comparing the k-means, fuzzy c-means and kernel fuzzy c means clustering algorithms for Davies–Bouldin index and Dunn index, the experimental results shows that kernel fuzzy c-means clustering algorithm which has a low Davies-Bouldin index and high Dunn index is the best algorithm. Hence kernel fuzzy c-means clustering algorithm is the best algorithm that can be applied on the images for evaluating the quality of the image based on internal criterion which gives the least value high computation complexity.

Fig. 9 shows the comparison chart for Davies-Bouldin index and Dunn index with four clusters and Fig. 10 shows the comparison chart for Davies-Bouldin index and Dunn index with six clusters. By analyzing, it can be clearly judged, that kernel fuzzy c means is the best among the k-means, FCM and KFCM clustering algorithms. KFCM which has low Davies-Bouldin index and high Dunn index is the most appropriate clustering algorithm for evaluating the quality of the image based on internal criterion.

4. CONCLUSION

There are various clustering algorithms available for multispectral satellite image segmentation. This paper proposes k-means, fuzzy c-means and kernel fuzzy c-means clustering algorithms. Davies–Bouldin index and Dunn index are applied on these algorithms to assess the quality clustering algorithms based on internal criterion. This paper focused the input image with four and six clustering regions extracted from the original image. The KFCM algorithm shows best results with low Davies–Bouldin index value and high Dunn index than the k-means and fuzzy c-means clustering algorithms. The KFCM clustering algorithm shows the computational complexity gets decreased due to the calculation of typicality degree but the effectiveness gets increased.

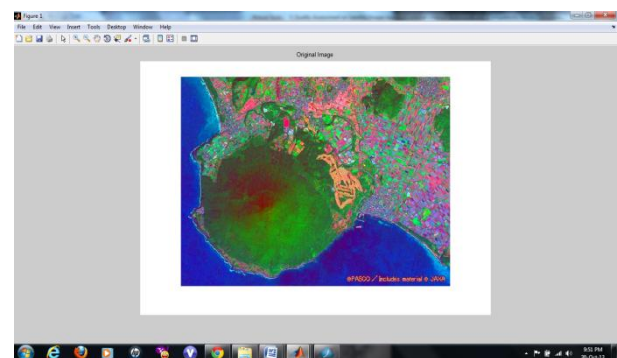


Fig. 1: Original image



Fig. 2: Image labeled by clustered Index

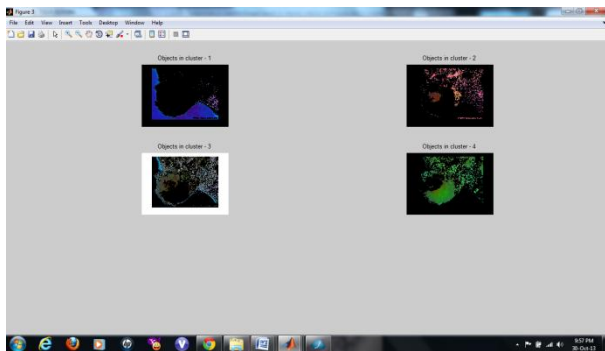


Fig. 3: Output of k-means algorithm with four clusters

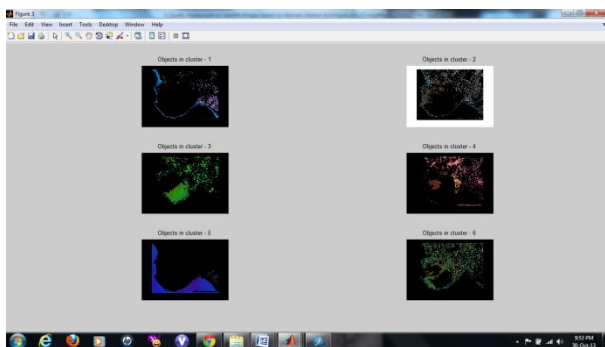


Fig. 4: Output of k-means algorithm with six clusters

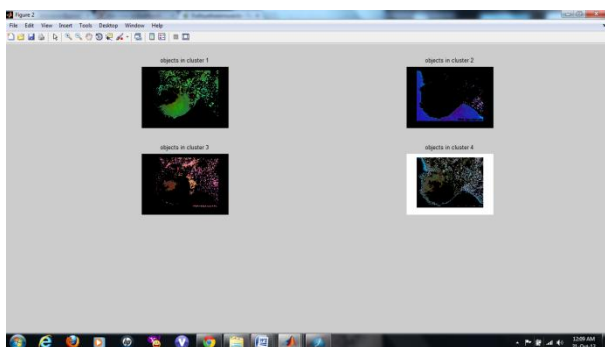


Fig. 5: Output of FCM algorithm with four clusters

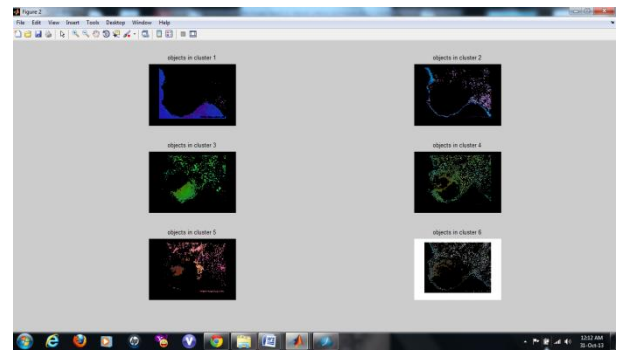


Fig. 6: Output of FCM algorithm with six clusters

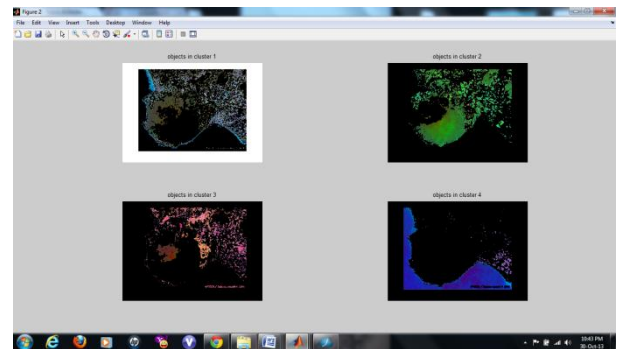


Fig. 7: Output of KFCM algorithm with four clusters

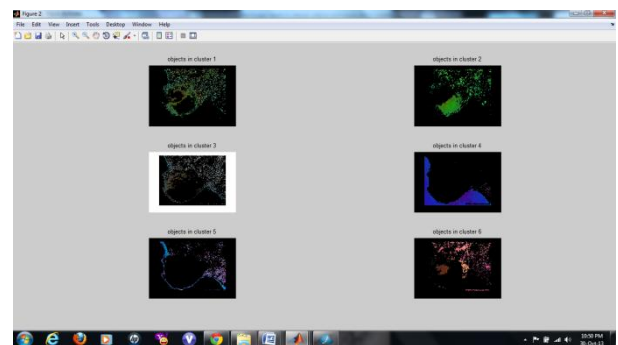


Fig. 8: Output of KFCM algorithm with six clusters

Table 1: Comparisons of the performance of different clustering algorithms for four cluster categories

Cluster Method	No. of clusters	Daveis-Bouldin Index	Dunn Index
k-means	4	0.6905	0.3182
FCM	4	0.6651	0.3725
KFCM	4	0.6328	0.4912

Table 2: Comparisons of the performance of different clustering algorithms for six cluster categories

Cluster Method	No. of clusters	Daveis-Bouldin Index	Dunn Index
k-means	6	0.7271	0.2531
FCM	6	0.6451	0.4083
KFCM	6	0.4275	0.6352

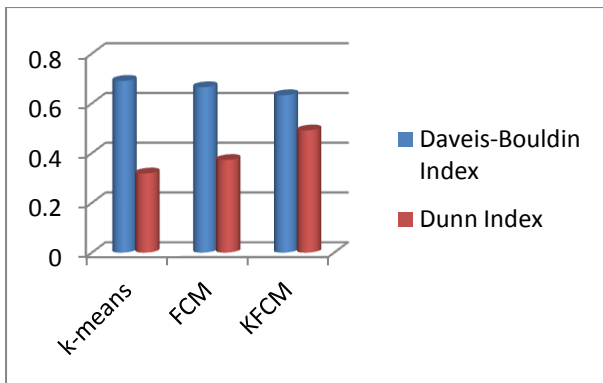


Fig. 9: Comparison chart for Davies-Bouldin index and Dunn index with four clusters

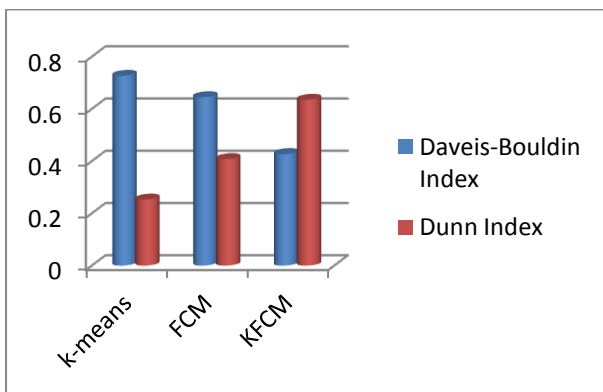


Fig. 10: Comparison chart for Davies-Bouldin index and Dunn index with six clusters

5. REFERENCES

- [1] Ahmed Darwish, et al, 2003. Image Segmentation for the Purpose Of Object-Based Classification, IEEE, pp 2039-2041
- [2] Anderberg M. R., 1973. Cluster Analysis for Applications. Academic Press.
- [3] Campbell, J. B., 2002. Introduction to Remote Sensing. New York London: The Guilford Press
- [4] Chen H C et al, 2006. Visible color difference Based quantitative evaluation of color segmentation, IEEE proceedings, Vis image signal process vol.153 No.5 pp 598-609.
- [5] Cheng, H. D, X. H. Jiang, Y. Sun and J. Wang, 2001. Color image segmentation: advances and prospects, Pattern Recognition, Vol.34, 2259 – 2281.
- [6] Coleman G. B. and H. C. Andrews, 1979. "Image segmentation by clustering", Proc IEEE 67:773–785.

- [7] Cristianini N, J.S. Taylor. An Introduction to SVMs and other kernel-based learning methods. Cambridge Univ. Press, 2000.
- [8] D.L. Davies and D.W. Bouldin, "A cluster separation measure ", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224-227, 1979.
- [9] Jain A. K., M. N. Murty and P. J. Flynn, 1999. "Data clustering: A review", *ACM Computing Surveys* 31 (3) 264–323.
- [10] Pratt W., 1991. Digital Image Processing, John Wiley & Sons, New York.
- [11] Zhang, Y. J. 2002. "Image engineering and related publications" *International Journal of Image and Graphics*, 2(3), 441- 452.

6. BIOGRAPHY

K. M. Sharavana Raju, Lecturer in Department of Computer Science, College of Computer Science and Information Systems, Jazan University, Jazan, Kingdom of Saudi Arabia. He received his M.Sc., (Physics), PGDCA and M.C.A., degrees from Bharathidasan University, Tiruchirapalli; M.Phil., (CS) degree from Periyar University, Salem and currently pursuing Ph.D., Computer Science at Bharathiar University, Coimbatore, Tamilnadu, India. He has published papers in 5 international journals, several national journals and also attended 3 International conferences and several national conferences and workshops. He is having 19 years of collegiate teaching experience, which includes 1 year in Libya. His main areas of research include Digital Image Processing, Data Mining and Signal Processing.

Mohammad Shahnawaz Nasir, Lecturer in College of Computer Science and Information System, Jazan University, Jazan, Kingdom of Saudi Arabia. He received his Master in Computer Science & Applications (MCA) and Master in Science (M.Sc.) Physics, Electronic Specializations Degrees from Aligarh Muslim University, India. Previously he worked as faculty with AOU (KSA), JMI (New-Delhi) and AMU (Aligarh, India). His work experience in Saudi Arabia also includes Web, Network, and Database Administration. Currently he is pursuing Ph.D., in Computer Science at Magadh University, Bodh-Gaya, India. His research areas are Data Mining, Biometrics and Image Processing.

Dr. T. Meera Devi, is a faculty member in the Department of Electronics and Communication Engineering of Kongu Engineering College, Perundurai. She received her Bachelor's degree in Electronics and Communication Engineering, in 1996 at The Indian Engineering College, Vadakkankulam, Tamilnadu, India; Master's degree in Applied Electronics in 2005 and Ph.D., in Signal Processing from Anna University, Chennai, Tamilnadu, India in 2013. She has been in the teaching profession for the past fourteen years. Her areas of academic interest include Soft Computing, Signal Processing and Image Processing.