

# Content based Image Retrieval for Medical Images using Canny Edge Detection Algorithm

B.Ramamurthy  
PhD Scholar  
PSG College of Technology  
Coimbatore 641004

K.R.Chandran  
Professor  
PSG College of Technology  
Coimbatore 641004

## ABSTRACT

The rapid expansion of digital data content has led to the need for rich descriptions and efficient Retrieval Tool. To develop this, content based image Retrieval method has played an important role in the field of image retrieval. This paper aims to provide an efficient medical image data Retrieval from a huge content of medical database using one of the images content such as image shape, because, efficient content-based image Retrieval in the medical domain is still a challenging problem. The main objective of this paper is to provide an efficient tool which is used for efficient medical image retrieval from a huge content of medical image database and which is used for further medical diagnosis purposes.

**Keywords:** Content Based Image Retrieval (CBIR), Support Vector Machine (SVM), Euclidean distance method(ED), Canny Edge detection Algorithm (CED).

## 1. INTRODUCTION

Content based image retrieval (CBIR) is a technique in which images are indexed by extracting their low level features and image retrieval is only based upon these indexed image features. [1, 2]. In an effective image retrieval system, the user poses a query and the system should find images that are somehow relevant to the query. Thus, a way of representing the query, a way of representing images, and a way of comparing a query and an image are needed. This kind of approach is known as querying by content. CBIR has accepted different kinds of user queries for its implementation such as query by example, query by color, and query by relevance. When a query image is given, the image is processed to extract features in order to represent the image contents as numeric values. These values are called feature vectors which are used in the retrieval process.

CBIR plays a pivotal role in medical image retrieval field such as CT scan images, MRI scan images, X-rays etc. Many systems already have been developed in this domain but each one has some difficulties in any way to provide efficient retrieval. So efficient content based image retrieval in medical domain is still a challenging problem. In this paper, we propose a method for medical image retrieval using canny edge detection algorithm. The following [Figure1] depicts a Classical architecture of CBIR system.

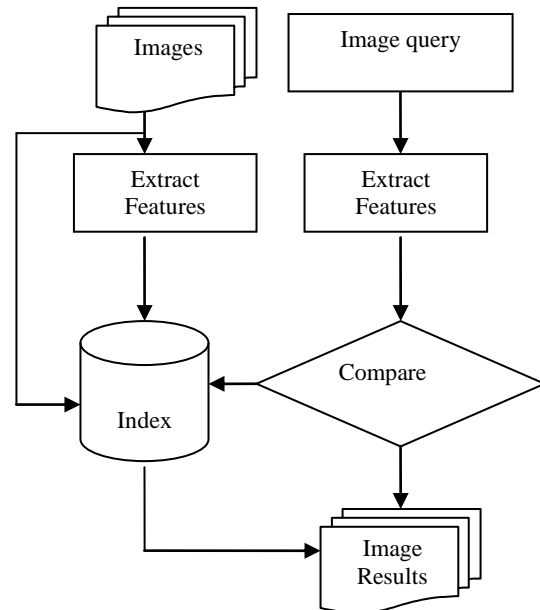


Figure1: Architecture of CBIR system

## 2. RELATED WORK

Medical images play a pivotal role in surgical planning, medical training, and patient diagnoses. In large hospitals thousands of images to be managed every year [3]. For images classifying, indexing and retrieval in manual method is very expensive one and time consuming because those medical images are vary person to person. In this section, we have reviewed some of existing works related to medical image searching techniques in content-based image retrieval.

The **Image Retrieval for Medical applications (IRMA)** project undertaken at the Aachen University of technology [4, 5] aims to provide visually rich image management through CBIR techniques applied to medical images using intensity distribution and texture measures taken globally over the entire image. This approach permits queries on a heterogeneous image collection and helps identify images that are similar with respect to global features. The IRMA system lacks the ability for finding particular pathology that may be localized in particular regions within the image.

The **Spine Pathology and Image Retrieval System (SPIRS)** [6, 7, 8] at the U.S. National Library of Medicine provides localized vertebral shape-based CBIR methods for pathologically sensitive retrieval of digitized spine x-rays and associated person metadata that come from the second U.S. National Health and Nutrition Examination Survey. In the SPIRS system, the images in the collection must be homogeneous.

The **Image Map** [9] is one of the existing medical image retrieval the considers how to handle multiple organs of interest. However, it works based on spatial similarity. Consequently, a problem caused by user is likely to occur and therefore, the retrieved image will represent an unexpected organ.

The **Automatic Search and Selection Engine with Retrieval Tools (ASSERT)** [10] A physician-in-the-Loop content-based retrieval system for HRCT image databases which is implemented to show a human-in-the-loop approach in which the human delineates the pathology bearing regions(PBR) and a set of anatomical landmarks in the image when the image is entered into the database.

The **MIMS** [11] takes the complexity first when describing semantic content of images and second the graphical aspect of certain objects. This existence creates some critical problems of subjectivity, although such an approach is made as general as possible not to specific one.

The **WebMIRS** [12] with this system, the user manipulates GUI tools to create a query. In response the system returns the values for given user query and displays the associated x-ray images.

All of these methods have their own advantages and disadvantages in their retrieval performance. In Image retrieval process, set of image features have been used for image retrieval. The following section describes different features set used by different authors in their medical image retrieval applications.

**Image features:** Image features are plays an important role in image retrieval processing. In order to perform image retrieval process, the extraction of suitable features from the images are very important and by which, both the query image and database images are compared to retrieval of very similar images to query image from the database. There are three level of feature extraction global, local and pixel. The simplest visual image features are directly based on the pixel values of the image. Images are scaled to a common size and compared using Euclidean distance and image distortion model [13].Local features are extracted from small sub-images from the original image. The global feature can be extracted to describe the whole image in an average fashion. The low-level features extracted from images and their local patches are color, texture, and shape [14].

#### **Color features**

The color feature is one of the most widely used visual features in image retrieval. Because the human vision system is more sensitive to color information than grey values of images [15].Generally color features are extracted using the color histogram technique [16].The color histogram describes the different colors distribution in an image in a simple and computationally efficient manner. Other color feature extraction techniques are region histogram, color coherence vector, color moments, correlation histogram etc.

#### **Texture features**

The texture feature is usually extracted by using filters method. The Gabor filter [17] is frequently used filter method in texture extraction. A variety of Gabor filters in different degree and their relative positions captures value at that specific frequency and direction. Texture can be extracted from this group of value distributions [3]. Other texture feature extraction methods are co-occurrence matrix, wavelet decomposition, Fourier filters, etc.

#### **Shape features**

Shape is an important and most powerful feature used for image classification, indexing and retrievals. Shape information extracted using histogram of edge detection. The edge information in the image is obtained by using the canny edge detection [18].Other techniques for shape feature extraction are elementary descriptor, Fourier descriptor, template matching, Quantized descriptors etc.

### **3. PROPOSED METHOD**

In this system, the input query image is given at runtime to get the features of the image. The processing of database images consist of three main stages namely image feature extraction, classification, and indexing that have been used in order to retrieve the similar images from the database. For retrieval process, Euclidean distance finding method has been used. The following processing steps have been applied in order to perform the retrieval of similar images from the database.

1. Medical images are given as input to the system.
2. For given input images, Shape features are extracted using canny edge detection method.
3. Based on the extracted shape features, image classification process has been performed using Support Vector Machine (SVM) tool.
4. Based on the SVM classification result, Indexing process has been performed using low-dimensional shape based indexing technique [19]
5. Finally, Searching and Retrieval process has been performed using one of the similarity measures such as Euclidean distance finding method.

#### **3.1. Feature Extraction**

The precision of image classification, image indexing and image retrieval mainly based on image feature extraction. More distinguished image features will yields better results in classification, indexing and retrieval process. In this work, we have extracted shape feature using canny edge detection algorithm [20]. Shape representations can be either edge or region based. Shape provides numerical information of an image, which do not change even when the position, size and direction of the objects are changed. By using canny edge operator, edge histograms of images are generated, which are given as input to the image classification tool (SVM) in order to classify the images. The following shows the canny edge detection algorithm steps. The algorithm runs in 5 separate steps.

1. **Smoothing:** Blurring of the image to remove noise.
2. **Finding gradients:** The edges should be marked where the gradients of the image have large magnitudes.
3. **Non-maximum suppression:** Only local maxima should be marked as edges.
4. **Double thresholding:** Potential edges are determined by thresholding.
5. **Edge tracking by hysteresis:** Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge.

The following [Figure 2(a) and 2(b)] depicts the results of before and after feature extraction of the work for two different kind of images.

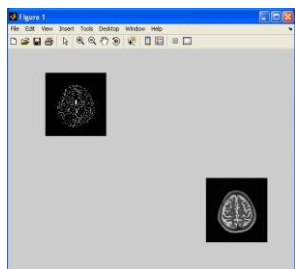


Figure2 (a): Before and After Feature extraction.

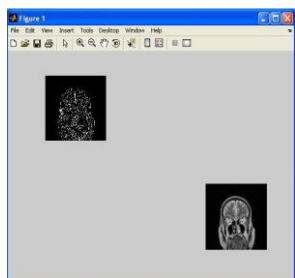


Figure2 (b): Before and After Feature extraction

### 3.2. Image Classification

Image classification is one of the important steps in image retrieval process because it saves more time while searching the images from huge volume of database. There are so many classifying techniques have been used to classifying the images such as neural networks, k-nearest neighbor pattern classifier, support vector machines (SVM) etc. Among which SVM have been shown to provide better generalization performance for many classification applications[21] than other techniques. In this work, we have used support vector machine (SVM) for classifying the database images in order to simplifying the searching process time in our database.

The following [Figure3] shows a typical support vector machines (SVM) classification.

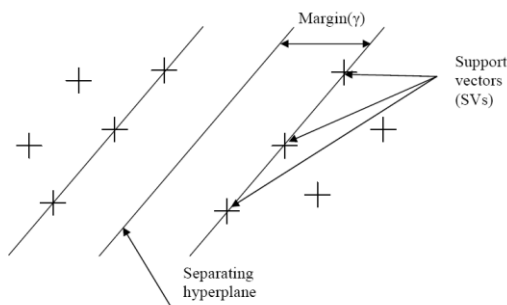


Figure3: Support Vector Machines Classification

In this step, a set of training samples given as input to the SVM in order to construct the binary classification of the images and each sample belongs to a class labeled. SVM use hyper planes in order to separate the two parts of the image classes such as positive, negative. This hyper plane causes the major separation among the decision function values for the two classes.

The following functions have to be performed in order to separate the classes of decision values. Let training samples are  $x_1, \dots, x_n$ , where, each sample belongs to a class labeled such as  $y_i \in \{+1, -1\}$ , then, the hyper plane decision function can be written as [22, 23]

$$f(x) = \text{sign} \left\{ \sum_{i=1}^n \alpha_i y_i \langle X_i, X \rangle + b \right\} \quad (1)$$

Where the coefficients  $\alpha_i$  and  $b$  are computed by quadratic programming problem.

$$\text{Max} \sum_{i=1}^l \alpha_i - 1/2 \sum_{i,j=1}^l \alpha_i \alpha_j \langle X_i, X_j \rangle Y_i Y_j \quad (2)$$

With

$$\sum_{i=1}^l \alpha_i Y_i = 0 \quad 0 < \alpha_i < C$$

Where  $C$  is regularization parameter.

The following [Figure4] shows mapping in support vector machine (SVM).

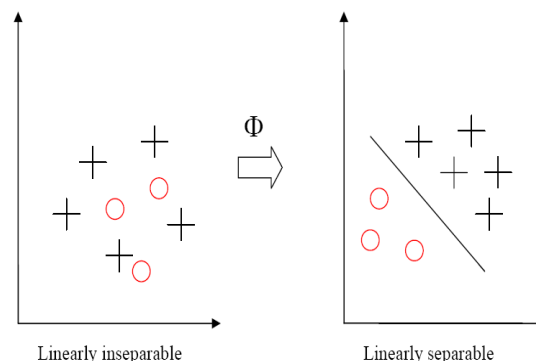


Figure4: Mapping in support vector machines (SVM).

### 3.3. Image Indexing

Content-based image retrieval (CBIR) aims at searching image databases for specific images that are similar to a given query image based on matching of features derived from the image content. In this work, we have focused on a low-dimensional shape based indexing technique for achieving efficient and effective retrieval performance. In this work, the shape features are extracted using the canny edge detection algorithm and then the features are grouped into two parts (positive, negative) using support vector machines (SVM). Then these feature groups are indexed using a spatial indexing method and then which are used for retrieval process.

### 3.4. Image Searching and Retrieval

There are various image retrieval methods that have been applied in various applications. One among which is Euclidean distance method for image similarity measurement [24]. In this work, we have used Euclidean distance method to retrieval of similar images from the database. When query image is given, the shape feature of that image is identified and which is compared with support vector machine classes such as positive and negative one by one by using Euclidean distance method [24]. Class (positive and negative) selection depends upon the query image's feature vector values. By comparing query image with support vector class images, the distance between images will be finding and shortest distance will be considered as best matching image in that matching process. Similarly, one after one best matching images will be finding by measuring the one after one shortest distance in order to retrieve the similar images from the database. The distance  $d$  between two image feature vectors is calculated by using the following equation.

$$d = \sqrt{\sum_{i=0}^N f_q((i) - f_d(i))^2} \quad (3)$$

Where,

$f_q(i)$  is the  $i^{\text{th}}$  feature vector of the query image

$f_d(i)$  is the  $i^{\text{th}}$  feature vector of the database image.

The following [Figure 5(a) and 5(b)] depicts retrieval results of the work for two different kinds of images.

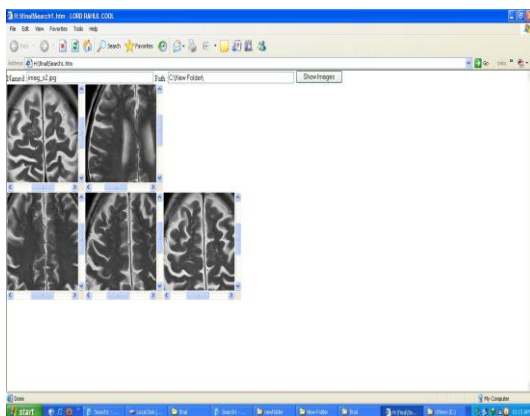


Figure5 (a): Retrieval result for chest image

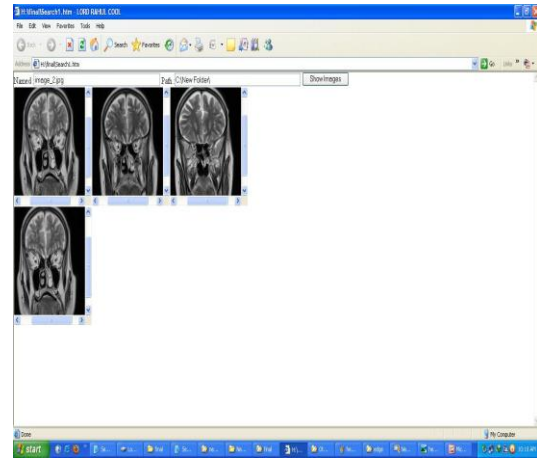


Figure5 (b): Retrieval result for brain image

## 4. EXPERIMENTAL AND RESULT ANALYSIS

We have collected 5000 real time images with 10 tests initially for testing the retrieval performance of the system. To measure this, precision and recall parameters were used and it was measured by using the following relationships

$$P=r/n1 \quad (4)$$

Where,

r-number of relevant images

n1-number of retrieved images

$$R=r/n2 \quad (5)$$

Where,

r-number of relevant images

n1-total number of relevant images in DB.

The following [Table1] and [Figure6] provides the precision and recall results recorded for various test images using CED shape detector and graph representation of the recorded results.

**Table1.Precision and Recall results recorded for various test images using CED shape detector**

Tests	Recall (%)	Precision (%)
1	15	60
2	25	58
3	35	55
4	45	52

5	55	48
6	65	43
7	75	35
8	85	20
9	95	12
10	100	10

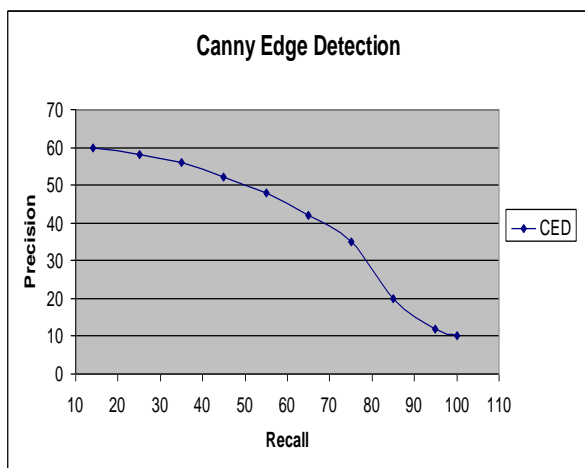


Figure 6. Precision and Recall graph using CED shape detector

Canny edge detection (CED) algorithm is one of the best algorithms since it provides more accurate edge detection values for applying searching algorithm in this work. By applying this algorithm, more accurate results were recorded as in the above given [Table1] test sets. Also this method will give more accurate results when scalability of the database is increased, because it provides more accurate edge detection values, since smoothing of the images in the process. In table1 given data shows the precision and recall values of the 10 data sets. When scalability of the database increased, the recall values of the database also increased [equation 5]. When recall values increased, the precision values are decreased [equation 4]. Due to the growth of database scalability in each test; recall values of the test sets are increased and precision values of the each test sets are decreased according to the growth of recall values. In this work, out of 100 recall values, about 10% of the precision values were produced. (ie 10:100 ratio). For 5000 images with 10 tests, about 500 relevant images were retrieved.

## 5. DISCUSSION AND CONCLUSION

In this work, we have used canny edge detection algorithm for extracting the shape features for the medical images. After extracting the shape feature, which are given as input to Support Vector Machine (SVM) for classifying the images and then classified images are indexed and labeled

for making easy for applying retrieval algorithm in order to retrieve the relevant images from the database. This work retrieves the images from the huge medical database as required by the doctors for their further evaluation about their patients' diseases. So that doctors can give proper treatment to their patients in proper time.

Due to the rapid expansion of medical image database, unfortunately, the scalability of the medical image database is also increased in day to day medical treatment activities. When the scalability of the medical image database is increased, then the performance of the retrieval process is decreased. Hence the performance of the retrieval process is needed to be increased more along with the large scalability of the database. To overcome this issue, the canny edge detection algorithm, Support Vector Machine (SVM), and spatial indexing method were used. This will give retrieval accuracy, speed, and clarity of the medical images to predict exact disease situation of the patients. So that doctors can get exact idea about disease in order to give the treatments for their patients.

Our future work is to provide symptoms description, remedies for the diseases along with medical images.

## 6. REFERENCES

- [1]. A Smeulder, M Worring, S Santini, A Gupta and R Jain, "Content based Image Retrieval at the end of the early Years", *IEEEPAMI*, vol22, no12, 2000, pp 1349.
- [2]. Y Rui, T S Huang and S F Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues", *Journal of Visual Communication and Image Representation*, vol 10, no1, 1999, p39.
- [3]. Muller, H., N.Michoux, D.Bandon and A.Geissbuhler, "A review of content based image retrieval systems in medical applications-clinical benefits and future directions", *International Journal of Medical Information.*, pp: 73:13, 2004.
- [4]. Lehmann TM, Guld MO, Thies C, Fischer B, Sphzer K, Keyzers D, et al, "Content-based image retrieval in medical applications", *Methods of Info in Med* 2004;73(1):1-23.
- [5]. Thies C, Guld MO, Fischer B, Lehmann TM, "Content-based queries on the CasImage database with in the IRMA framework", *Lec Notes in Comp Sci* 2005; 3491:781-92.
- [6]. Antani S, Long LR, Thoma GR, "Content-based Image retrieval for large biomedical image archives. In: Proc 11<sup>th</sup> World Cong Medical Informatics: 2004. P.829-33.
- [7]. Long LR, Antani SK, Thoma GR, "Image Informatics at a national research center, *CompMed Imaging &graphics* 2005; 29:171-93
- [8]. Thoma GR, Long LR, Antani SK, Biomedical Imaging research and development: knowledge from images in the medical enterprise. Technical Report Lister Hill National Ctr for Biomedical Communications, US National Library of Medicine, NIH 2006; LHCNBC-TR-2006-002.
- [9]. Petrakis, Euripidies G.M and C.Faloutsos, "ImageMap: An Image indexing Method Based on Spatial Similarity",

- IEEE Trans on Knowledge and Data Eng., 14(5):979-987, 2002.
- [10]. Chi-Ren Shyu, Carla E. Brodley, Avinash C. Kak, Akio Kosaka, "ASSERT: A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases", *Computer Vision and Image Understanding*, Vol.75, Nos.1/2, July/August, pp.111-132, 1999.
- [11]. Chbeir, R., Y. Amghar, and A. Flory, "MIMS: A Prototype for Medical Image Retrieval", Proctor the 6<sup>th</sup> Conference on Content-Based Multimedia Information Access- Recherché Informations Assistée par Ordinateur. Paris, France, 2000.
- [12]. Long LR, Pillemer SR, Lawrence RC, Goh GH, Neve L, Thoma GR. WebMIRS: Web-based Medical Information Retrieval System. Proceedings of SPIE Storage and Retrieval for Image and Video Databases VI, SPIE vol.3312, San Jose, CA, January 24-30, 1998. pp. 392-403.
- [13]. Keysers, D., C. Gollan and H. Ney, "Classification of medical images using Non-linear distortion models", in: *Bildverarbeitung für die Medizin*, Berlin, Germany, 2004.
- [14]. Ma, W.Y. and H.J. Zhang, "Content Based Image Indexing and Retrieval". In: *Handbook of Multimedia Computing*, (Ed.) Furht, B. CRC Press, Boca Raton, 1998.
- [15]. Lim, J.H., S.J. Jesse and Luo Suhuai, "A Structured Learning Approach to Semantic Photo Indexing and Query", *Asia Information retrieval symposium*, 13-15 October 2005, Jeju Island, Korea.
- [16]. Swain, M.J. and D.H. Ballard, "Color Indexing", *International Journal of Computer Vision*, 7:11-32, 1991
- [17]. Turner, M.R., 1986. Texture discrimination by Gabor functions. *Cybernetics*, 55:71-82.
- [18]. Canny, J., "A computational approach to edge detection." *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:679-698, 1986.
- [19]. J Z Wang, Gio Wiederhold, O Firscheing and S X Wei. "Content-Based image Indexing and Searching using Daubechies Wavelets" *International Journal on Digital Libraries*, vol 1, Jan 1998.
- [20]. J F Canny." A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 8, Nov 1986.
- [21]. Wang, Y., B.C., Long and G.H. Bao, "Semantic extraction of the building images using support vector machines". In *Proceedings of 1<sup>st</sup> International Conference on Machine Learning and Cybernetics*, November 2002, pp: 1608-1613
- [22]. Vivek Balasubramanyam<sup>1</sup>, Andreas H. Hielscher, "Classification of Optical Tomographic Images of Rheumatoid Finger Joints with Support Vector Machines" *Proc. of SPIE Vol. 5692*, (SPIE, Bellingham, WA, 2005) pp:37-43.
- [23]. A. Mueen, M. Sapiyan Baba and R. Zainuddin, "Multilevel Feature Extraction and X-ray Image Classification" *Journal of Applied Sciences* 7(8):1224-1229, 2007.
- [24]. A. Grace Selvarani and S. Annadurai, "Content Based Medical Image Retrieval System using Shape and Texture Features", *ICGST-BIME Journal*, Vol 8, Issue 1, December 2008.

## 7. AUTHORS PROFILE

**B. Ramamurthy** received his UG and PG from Bharathidhasan University, Trichy, and M.Phil. Degree from Bharathiar University, Coimbatore. He has 12 years of teaching experience. He is currently working as Senior Lecturer in the Department of Computer Science at Sri Ramakrishna Engineering College, Coimbatore, India. He is a member in ACS. His areas of research include Digital Image processing for image retrieval, Multimedia, and Software Engineering. Currently he is pursuing Ph.D., at PSG College of Technology, Coimbatore.

**K.R. Chandran** received his BE in 1982 and ME in 1986 from PSG College of Technology and PhD in 2006 from Bharathiar University, Coimbatore. He is currently professor and Head in the Department of Information Technology at PSG College of Technology, Coimbatore. He is a Fellow member in Institution of Engineers (India) and life member of ISTE, ACS, and The Textile Association (INDIA). He has published more than 50 research papers in various journals and conferences. His research interest includes software Engineering, Computer Networks, Data Mining, System Modelling and Simulation.