

Lung Segmentation and Iterative Weighted Averaging Smoothing Technique on Chest Ct Images



C. Sridevi, M.Kannan

Abstract: *Computed Tomography (CT) is one of the most commonly used imaging modalities for tumour detection and diagnosis, due to its high spatial resolution, fast imaging speed and wide availability. Nodules of the lungs and pathological residues with varied diameter can be comfortably viewed by computed tomography and can be categorized as benign or malignant. The key intention of this segmentation and smoothing is to develop an efficient methodology for an automated lung tumour diagnosis. Segmentation is the quantitative pre-processing step used in the chest CT analysis. The iterative weighted averaging technique is used in addressing the issues related to the segmentation and smoothing method employed in this paper. The methodology incorporates the accurate Lung Segmentation, enclosure of Juxtapleural nodules, the proper insertion of the bronchial vessels for enhancing the smoothness of the segmented Lung area. The steps involved in this paper include Image preprocessing by nonlinear anisotropic diffusion filtering, Thorax Extraction, Lung segmentation and iterative weighted averaging technique for smoothing the contours. The main feature in choosing the nonlinear anisotropic diffusion filtering is for properly preserving the regions unaffected with cancer and also to differentiate the artefacts involved due to the subject movements. In the fuzzy c- means clustering algorithm, the lung parenchyma is identified from the thorax region based on the fuzzy membership value and the affected regions are attached. The standard execution time for segmentation process of one dataset is 1.91s, it is the more rapid process than the manual segmentation.*

Keywords: Lung Segmentation, Lung Nodules, Fuzzy c-means Clustering, Iterative Weighted Averaging.

I. INTRODUCTION

Lung cancer is the uncontrolled growth of abnormal cells in one or both lungs. These abnormal cells do not carry out the functions of normal lung cells and do not develop into healthy lung tissue. Lung cancer leads the rate [9], over 1.5 Lakh of men and women patients are died per year including the cervical, breast, prostate cancer in the United States were getting affected, according to the survey cited by American Cancer Society. The lung nodule identification is done from the computed tomography images (CT) for the early

detection of the cancer cells in the subject and to increase their survival rate.

The main objective of segmentation in the medical images is the acquisition of proper boundary of the lung structure for the clinical analysis and also for quantitative data analysis, for identifying the severity of the condition. The procedures of the algorithms are varied because of the properties such as shape, size, intensity and location of the cancerous cells. The analysis of data 6- 16% of the nodules are inaccurate and marked wrongly because of the choice of inappropriate segmentation algorithms.

As a result, the design of an automated, reliable, and accurate computerized segmentation tool is used to identify the Cancer cells in the Lungs. Thresholding method pooled with region growing [11, 12] is generally used to extract lung parenchyma, which is usually expelled from the thoracic region because of its higher gray level. To overcome this issue of loss of juxtapleural nodules and to obtain the smoothness of the lung boundary, many lung contour algorithms are proposed. In [13] the loss of the nodules is avoided by using proper threshold values for segmentation process. The extraction of nodules is the most challenging task because of its size variation and can be properly identified using the rolling ball algorithm [7]. Bellotti et al. [12] designed a new active contour, called glued elastic band, to correct the contour of lung parenchyma, which rely on the predefined threshold, unfortunately, it cannot be chosen automatically but empirically. Pu et al [13] suggested an ABM algorithm to extract a proper lung boundary for the cancer cell analysis. The major drawback of this algorithm is the usage of threshold values which may lack proper identification of the borders.

Varshini et al. [6] suggested a algorithm with two different threshold values to solve the problem. In this paper, the segmentation and correcting the lung boundary is done by segmentation by registration scheme. This algorithm is applied on the images of lungs with severe abnormalities, but the algorithm encounters inconsequential issues due to the higher degree of pathological severity in Lungs.

In addition to the above mentioned survey, some of the literature also discusses [15] about extracting the correct boundary of the lung extraction along with the nodules and the bronchiole vessels. In most of the literature reviews the smoothness of the lung contour and the perfect segmentation of the same is of a challenge.

Manuscript published on 30 December 2019.

* Correspondence Author (s)

C. Sridevi*, Department of Electronics Engineering, Anna University, Chennai, India. Email: sridevijegan@annauniv.edu

Dr. M. Kannan, Department of Electronics Engineering, Anna University, Chennai, India. Email: mkannan@mitindia.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

II. METHODOLOGY

The flow diagram of the work is illustrated in fig 1. The first step of the procedure involves the acquisition of the computed tomography images from the database. The preprocessing of the images is done by the nonlinear anisotropic filtering technique. The initial lung boundary is obtained by the fuzzy c means algorithm after the region growing process. The segmented smooth contour of the lung boundary is identified and extracted using the Iterative weighted averaging method.

General steps to be followed during implementation of image pre- processing and image segmentation illustrated below:

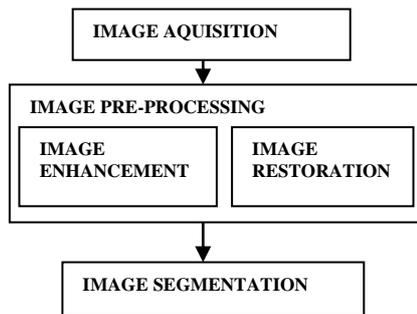


Figure 1: General steps of Image Pre-processing and Segmentation

III. IMAGE DATA SET

The main benefit of using computed tomography images is less distortion. CT images have low noise in contrast to X-ray and MRI images; henceforth they are used most commonly in the Medical image analysis. The CT images are acquired from the standard database for Medical imaging archive Lung Image Database Consortium (LIDC) dataset, DICOM (Digital Imaging and Communications in Medicine). The DICOM images are converted to JPG format for ease of processing. The acquired images are in raw form and have a lot of noise, thus to improve the contrast, clarity and to separate the background noise, it is required to pre-process the images. Thus, different methods like smoothening, segmentation are applied to get image in required form. Also taken some image samples from the Cancer Imaging Archive TCIA Collection.

IV. IMAGE PREPROCESSING

The idea of pre-processing minimizes the noise present in the image and helps in minimizing the complexity of the steps involved in the algorithm. The non linear anisotropic filter reduces the problems related to blurring of edges and losing of important details present in the image for further processing. The filter also enables to preserve the region information, edges, lines and minimizes the loss after the filtration process. It also produces better results compared to the other types of most popularly used filters for image processing applications.

$$\frac{\partial I}{\partial t} = \text{div}(c(\|\nabla I\|)\nabla I) \quad (1)$$

Where $\|\cdot\|$ and div denote the L2 norm and the divergence operator of the associated quantities, respectively; ∇ represents the gradient of the diffusing image I ; and $c(\cdot)$ is a

diffusivity function, also referred to as the diffusion coefficients, function of the modulus of the image intensity gradient.

V. HISTOGRAM EQUALIZATION

Histogram equalization adjusts the contrast of the image using the histogram representation. The image obtained is the equalized image of close contrast values by keeping the global contrast values as reference. In histogram equalization, the images with both foreground and background dark or bright are used. For equalizing the histogram, we need to compute the histogram and then normalize it into a probability distribution. For normalization, we just need to divide the frequency of each pixel intensity value by the total number of pixels present in the image. This is equal to the resolution of the image. i.e. $(N_{rows} \times N_{cols})$. The equalization process makes sure that the resulting histogram is flat. The cumulative distribution function (cdf) is given by:

$$cdf(x) = \sum_{k=-\infty}^x P(k)$$

The main initiative of this step is to obtain the resultant image a linear cumulative distribution function, where the intensities are evenly distributed on the histogram. In the Histogram plot (Fig. 3), the pixel intensity is plotted in the y-axis and the frequency of occurrence of the Pixel values is mentioned in the x-axis.

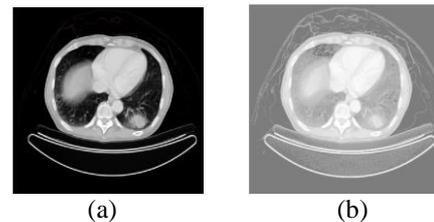


Figure 2: (a) Original Image, (b) Histogram Equalized Image

VI. THORAX EXTRACTION AND LUNG IDENTIFICATION

This step involves the extraction of the thorax region from the others by the process of thresholding. This procedure is complicated because of the presence of muscle structures which hinders the process to be successful. This stage of extraction is considered to be very mandate in the lung segmentation in order to minimize the computational cost in the further steps involved in the procedure. The procedure can be taken up in three steps: outer layer of the lungs are identified by the fuzzy c- means method, then most muscular regions and bronchiole structures are removed using region growing algorithm and also the lung regions may be connected for the analysis purpose.

VII. REGION GROWING

It is a region based image segmentation methodology where the regions are segmented based on the selection of the initial seed points.

In this segmentation method the similarity constraints are initially set by iteratively merging the set of small areas. The procedure involves growing of regions from the seed pixel and appending the pixels possessing similar characteristics, thereby expanding the size of the region to be segmented. The above procedure is repeated for forming different regions with different seed points.

The similar regions identified based on the information obtained from the histogram of the image, threshold value, criterion of region growing and the pixel intensity, which determine the appropriate threshold value for the region membership criterion.

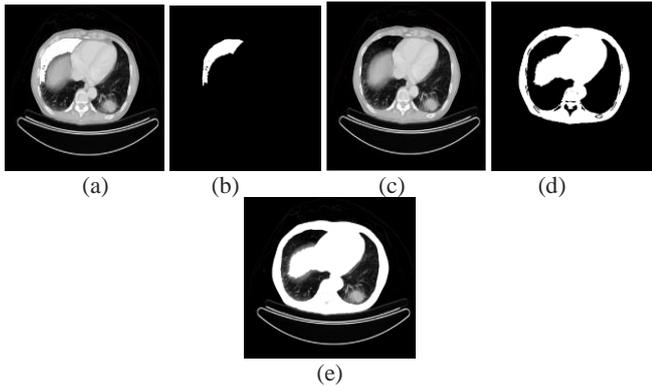


Figure 4: (a) & (b) During region growing (gathering seed points), (c) Original image, (d) Region Growing, (e) Concatenated Image after Region growing

VIII. FUZZY C-MEAN CLUSTERING ALGORITHM

The algorithm is an iterative clustering method that produces an optimal membership value by minimizing the weighted within group sum of squared error objective function obtained from the equation (2). The algorithm selects the number of clusters, assigns coefficients randomly [4] to the data points in the clusters, till the algorithm converges i.e. the coefficients changes between the two iterations within the given threshold.

The above mentioned clustering algorithm can be given as the minimization of the weighted inter-class sum of the squared error objective function.

$$J_{FCM} = \sum_{c=0}^C \sum_{j=0}^J u_{jc}^2 \|I_j - V_c\|^2 \quad (2)$$

where V_c is the centroid of class c and C – total number of tissue class; u_{jc} - membership value of pixel j for class c and requires $u_{jc} \in [0, 1]$ subject to, $\sum_{c=1}^C u_{jc} = 1$; J - total number pixels in the image and I_j - image intensity at the position.

$$u_{jc} = \frac{\|I_j - V_c\|^{-2}}{\sum_{l=1}^J \|I_j - V_l\|^{-2}} \quad (3)$$

$$V_c = \frac{\sum_{j=1}^J u_{jc}^2 I_j}{\sum_{j=1}^J u_{jc}^2} \quad (4)$$

The required objective function is converged, when the upper limit value is assigned to the pixel near to centroid and the lower limit value is assigned to pixel away from the centroid. The algorithm also solves the non-linear problems by continuous iteration of different parameter values. After every consecutive iteration, pair of membership functions and centroids is generated. The algorithm is as follows and the simulated output of the proposed algorithm is depicted in the Fig.5 (a) & (b).

Algorithm

- (a) Set the seed points for class centroids, $v_c, c = 1, 2, \dots, C$.
- (b) Calculate the parameter u_{jc} using (3).
- (c) Calculate the V_c using the equation (4).
- (d) Re-iterate the steps (b) and (c) till the process get saturated and the values obtained using equation (3) and (4) does exceed the threshold value of 0.001.

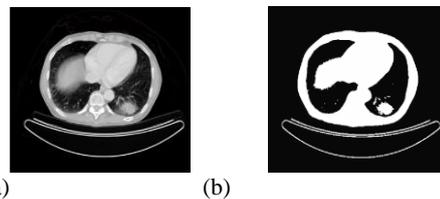


Figure 5: (a) Original image, (b) Fuzzy C means Clustering

The membership plot of the fuzzy c-means is illustrated in the Fig. 6

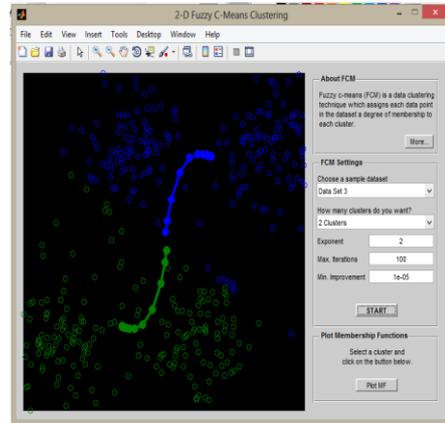


Figure 6: Membership plot for Fuzzy C Means Clustering

IX. ITERATIVE WEIGHTED AVERAGING

Iterative weighted averaging algorithm [14] smooth the image by finding the average of the neighborhood pixels. Let $P_i^{(0)} = (x_i, y_i)$ be a two-dimensional image at the point i before smoothing. The smoothed signal $P_i^{(n_2+1)}$ on $(2s+1)$ neighbourhoods at the $(n_2+ 1)$ th iteration is

$$P_i^{(n_2+1)} = \sum_{j=-s}^{j=s} S_{i+1}^{(n_2)} d_{j+s} \quad (5)$$

with

$$\sum_{j=-s}^{j=s} g_{j+s} = 1 \quad (6)$$

The coefficients d_i are set symmetrically comparative to the centre, that is $d_{s+j} = d_{s-j}$ ($j = -s, \dots, s$).

The expression (6) is corresponding to $\sum_{j=0}^{2s} g_j = 1$

The selection of g_j ($j = 0, \dots, 2s$) is supple, because it satisfies the above equations.

The value of s is set as 1, in order to extend the algorithm to the nearest-neighborhood pixels.

$$P_i^{(n_2+1)} = P_{i-1}^{(n_2)} d_0 + P_i^{(n_2)} d_1 + P_{i+1}^{(n_2)} d_2 \quad (7)$$

Where $d_0 + d_1 + d_2 = 1$. Taking $d_0 = d_2 = q$ and $d_1 = 1-2q$, $q =$ constant and can be minimized to

$$P_i^{(n_2+1)} = q(P_{i-1}^{(n_2)} + P_{i+1}^{(n_2)}) + (1-2q) P_i^{(n_2)} \quad (8)$$

Simplifying and ordering we get

$$P_i^{(n_2+1)} - P_i^{(n_2)} = q(P_{i-1}^{(n_2)} - 2P_i^{(n_2)} + P_{i+1}^{(n_2)}) \quad (9)$$

Equation (9) resembles the linear diffusion equation of discrete approximation $\frac{\partial s}{\partial t} = q \nabla^2 s$

Here, d_i is chosen as Newton–Cotes coefficient in the mathematical subfield of numerical analysis since it satisfy the said equations previously. The expression of Newton–Cotes coefficient d_i with the equal sample size w can be formulated as

$$d_i = \frac{(-1)^{r-i}}{ri!(m-i)!} \int_0^r \prod_{\substack{k=0 \\ k \neq i}}^r (t-k) dt, i = 0, \dots, 2s \quad (10)$$

Where $r = 2s$. The value of d_i is not required if P_i is sampled at equal step size which helps the algorithm to be less complicated.

After this sampling procedure the lung contour region can be tracked automatically with the edge tracking algorithm which is used to identify the closed initial contour. Starting from the i^{th} edge point on the initial contour at $P_i \in R^2$, pixel points are saved in anti- clockwise direction.

Given in $P_i^{(n_2)}$ the n_2 th iteration for point i , the $(n+1)$ th iteration based on three neighbours can be given below:

$$P_i^{(n_2+1)} = \frac{1}{6} P_{i-1}^{(n_2)} + \frac{2}{3} P_i^{(n_2)} + \frac{1}{6} P_{i+1}^{(n_2)} \quad (11)$$

The Fig. 7 (a) and (b) shows the results obtained using the iterative weighted averaging methodology on the lung contour for extracting a well smoothed contour of the segmented Lung area.

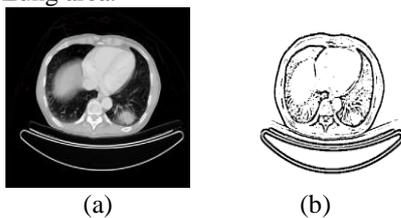


Figure 7: (a) Original Image (b) Iterative Weighted Averaging

I. Comparative Study

The output of fuzzy C means method is compared with the outputs of two different algorithms to demonstrate that FCM is better over other hard segmentation techniques. The algorithms used for comparison are a) Watershed algorithm b) Otsu’s method. In the Watershed Algorithm, it detects the Large nodules accurately and the implementation rate is high whereas the small nodules are eliminated and accuracy is less in presence of noise. In the Otsu's method, the smaller as well as larger circumscribed nodules are segmented accurately whereas the airways in the lung parenchyma region are not eliminated accurately. The algorithm gives better result in salt and pepper noise but not on Gaussian noise. The processing time for each step of our method is given below.

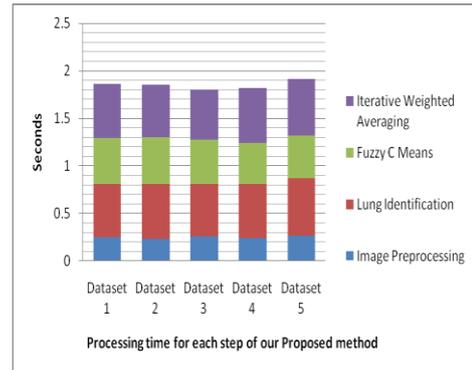


Figure 8: Processing time of the proposed method

The table 1 displays the advantages and disadvantages of the proposed method from the other two methods.

X. CONCLUSION

The CT images are pre-processed to trim down the noise and the artefacts due to patient movement using nonlinear anisotropic diffusion filtering. By employing histogram equalization technique the contrast of the image is improved by effectively spreading out the most frequent intensity values. By adapting region growing, the initial seed points are generated by examining the neighboring pixels and concluded as how the neighbor pixels are appended to the region of interest. We also discussed an algorithm for lung segmentation which includes juxtapleural nodules and the bronchial vessels of the thorax region. This algorithm enables to obtain a optimum method for obtaining a smoothed lung boundary. The initial lung boundary is extracted by the fuzzy c- means algorithm. To extract an accurate smooth lung contour in early stage of the cancer by averaging the neighborhood pixels, the iterative weighted averaging methodology is employed. Segmentation of Lung contours in case of severe pathological condition is still a challenging task and can be focused in the future work.

ACKNOWLEDGEMENTS

The first author C.Sridevi is receiving Visvesvaraya PhD scheme Fellowship for Electronics and IT from DeitY, Government of India (Lr. No. 3408/ PD6/ Deity/ 2015 dated 30-06-2017) for her research work.



Table- I

METHODS	FCM	WATERSHED ALGORITHM	OTSU'S METHOD
ADVANTAGE	Detects the circumscribed nodules and eliminates the airways in the lung parenchyma. Higher accuracy.	Detects the large sized nodules. Faster implementation.	Detects the smaller as well as larger circumscribed nodules.
DISADVANTAGE	Pulmonary and juxtaleural nodules are not included in the output.	Smaller nodules are eliminated and accuracy is less in presence of noise.	It does not eliminate the airways in the lung parenchyma region. Less accurate.
IMAGE OUTPUTS			

REFERENCES

- J. Jeya Caleb, M. Kannan, Design of an Improved K-Means Algorithm for the Clustering Of Data, Journal of Chemical and Pharmaceutical Sciences, Volume 10 Issue 3, ISSN: 0974-2115, July – September 2017.
- J. Jeya Caleb, M. Kannan, VLSI Implementation of Constructive Neural Network for Skin Cancer Detection, Journal of Computational and Theoretical Nanoscience, Vol. 15, 485–492, 2018.
- Ratishchandra Huidrom, Yambem Jina Chanu, Khumanthem Manglem Singh, A Fast Automated Lung Segmentation Method for the Diagnosis of Lung Cancer, Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017, 1499 - 1502.
- Bhagyarekha U. Dhaware, Anjali C. Pise, Lung Cancer Detection Using Bayasein Classifier and FCM Segmentation, 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 978-1-5090-2080-5/16, 2016 170 -174.
- Shengjun Zhou, Yuanzhi Cheng, Shinichi Tamura, Automated lung segmentation and smoothing techniques for inclusion of juxtaleural nodules and pulmonary vessels on chest CT images, journal on Biomedical Signal Processing and Control, Volume 13 , Pages 62-70, September 2014.
- P.R. Varshini, S. Baskar, S. Alagappan, An improved adaptive border marching algorithm for inclusion of juxtaleural nodule in lung segmentation of CT-images, IEEE Int. Conf. Image Process. (2012) 230–235.
- S. Sivakumar, C. Chandrasekar, Lungs image segmentation through weighted FCM, in: IEEE International Conference on Recent Advances in Computing and Software Systems, 2012, pp. 109–113.
- Q. Li, F. Li, K. Suzuki, J. Shiraishi, H. Abe, R. Engelmann, Y. Nie, H. MacMahon, K. Doi, Computer-aided diagnosis in thoracic CT, Seminars in Ultrasound, CT, MRI 26 (5) (2005) 357–363.
- Cancer Facts & Figures, American Cancer Society: Cancer Statistics, 2012 <http://www.cancer.org>.
- C.I. Henschke, D.I. McCauley, D.F. Yankelevitz, D.P. Naidich, G. McGuinness, O.S.Miettinen, D.M. Libby, M.W. Pasmantier, J. Koizumi, N.K. Altorki, J.P. Smith, Early lung cancer action project: overall design and findings from baseline screening, Lancet 354 (9173) (1999) 99–105.
- S.G. Armato, W.F. Sensakovic, Automated lung segmentation for thoracic CT: impact on computer-aided diagnosis, Acad. Radiol. 11 (9) (2004) 1011–1021.
- R. Bellotti, F. De Carlo, G. Gargano, S. Tangaro, D. Cascio, E. Catanzariti, P. Cerello, S.C. Cheran, P. Delogu, I. De Mitri, C. Fulcheri, D. Grosso, A. Retico, S. Squarcia, E. Tommasi, B. Golosio, A CAD system for nodule detection in low-dose lung CTs based on region growing and a new active contour model, Med. Phys. 34 (12) (2007) 4901–4910.
- J. Pu, J. Roos, C.A. Yi, S. Napel, G.D. Rubin, D.S. Paik, Adaptive border marching algorithm: automatic lung segmentation on chest CT images, Comput. Med. Imaging Graph. 32 (6) (2008) 452–462.
- E.M. Van Rikxoort, B. de Hoop, M.A. Viergever, M. Prokop, B. van Ginneken, Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection, Med. Phys. 36 (7) (2009) 2934–2947.
- Y. Yim, H. Hong, J. BeomSeo, N. Kim, E. Jin Chae, Y. Gil Shin, Correction of lung boundary using the gradient and intensity distribution, Comput. Biol. Med. 39(3) (2009) 239–250.
- [16] Shakouri G., Hamed & B. Menhaj, M & Moradmam, Hashem, Introduction to an Iterative Weighted Mean Smoothing Filter Based on a Simple Fuzzy Rule. Journal of Amirkabir University of Technology. 2006. 27-33.

AUTHORS PROFILE



C Sridevi graduated from the Department of Electronics and Instrumentation Engineering, Bharathiyar University, India in 2001 and completed M.E. degree in Applied Electronics from Anna University, India in 2005. She is employed as Assistant Professor in the Department of Electronics Engineering, Madras Institute of Technology campus of Anna University, India. Her areas of interests include Image Processing, Biomedical Image Processing, Parallel Architectures and Neural Networks. She has authored and co-authored many papers in these fields. She is a Fellow member of IEEE , Life member of IAENG.



Dr.M Kannan received his B.E degree in ECE from Alagappa Chettiar Government College of Engineering & Technology, 1988 and M.Tech degree in Electronics from Anna University in 1995. He received his Ph.D. in the field of Digital VLSI Design from Anna University in 2007. He is serving as Professor, Department of Electronics Engineering, Madras Institute of Technology, Anna University, Chennai. His main research interest includes Digital VLSI design, Computer Architecture and Computer Networks. He has published more than 75 papers in national/international journals and conferences. He is a Fellow member of IEEE, Life member of ISTE, ACCS and CSI.

