A NON INVASIVE COMPUTER AIDED DIAGNOSIS SYSTEM FOR EARLY DETECTION OF LUNG CARCINOMA IN CT MEDICAL IMAGES

V.I.Mebin Jose *(Research Scholar)*[1], Dr.T.Arumuga Maria Devi *(Assistant Professor)*[2]

***Abstract-*Lung cancer is a potentially fatal disease caused mainly by environmental factors that mutate genes encoding critical cell regularities proteins. This paper investigates the early detection of lung cancer to improve the performance of computer aided diagnosis system. The Existing multimodal sparse representation based classification cannot produce the optimal detection rate due to the in consistency of feature selection. Moreover, the computation time is also high. To overcome these limitations, a novel cancer detector developed with higher detection rate with minimal computation time is called EHFD. The proposed (EHFD) Enhanced and Hybrid Feature based Detector outperforms well in real time because of combining optimal features from enhanced and non-enhanced images. The experimental results validated that the proposed method improved the IEF by 19% and detection overshoot by 93% compared with existing methods. Therefore, the proposed method is most suitable for early estimation and diagnosis of lung cancer pretty well. The produced results demonstrate that the detection performance fits for diagnosis of lung cancer in early stage and integrate with the current diagnosis system.**

Keywords—Enhanced and Hybrid based Feature Detector, Image Enhancement Factor, Local binary Pattern, stationary wavelet transform

I. INTRODUCTION

According to the report from the World Health Organization, lung cancer is the most common cause of cancer- related death worldwide with an estimated 27% of all cancer deaths. it is the second most common cancer in both men and women along this Non-small cell lung cancer (NSCLC) is mostly affected cancer compared to other kind of lung cancers. The leading cause of cancer is death. Every year, many people died because of lung cancer other than colon, breast, and prostate cancers combined. The overall five-year relative survival rate of lung cancer for all stages is approximately 15%, but if the primary tumor is treatable, the survival rates are improved [2]. Therefore, accurate staging is critical for prognosis and appropriate management. Non-small cell lung cancer (NSCLC) is the major type of lung cancer.

Conventionally, lung cancer can be categorized into four types: squamous carcinoma (SC), adenocarcinoma (AC), small cell cancer (SCC), and Nuclear Atypia (NA). Non-small cell lung cancer (NSCLC) is more predominant when compared to varied types of lung cancer. The key to lung cancer treatment relies on its early stage diagnosis only. Traditionally, the diagnosis of lung cancer is made by the pathologist. However, the pathologist's inexperience in clinical practice can cause misdiagnosis. Eventhough we have so many records to prove how deadly lung cancer can be, still nothing is done to increase the survival rate of Non small cell lung cancer and the research is still slow. The two main reasons behind this is that very little federal funding is commited to lung cancer research and very few cases are diagnosed at early stage when cancer is most curable. Hence centre should plan and excecute fund raisal especially in the lung cancer research which will definitely prove to be effective in the longer run

The worst part is that the diagnosing process is time consuming and sometimes tedious when a pathologist is asked to analyze a huge volume of sample images from patients. The accuracy of the diagnosis is related not only to the pathologist's educational background and clinical experience but also his/her physical and psychological conditions. Thus, misdiagnosis of cancer can happen even for an experienced pathologist. Compared to traditional manual reader approaches, computer-aided approaches offer numerous advantages such as reducing inter and intra-reader variability. The best imaging technique to stage NSCLC is the combined positron emission tomography (PET)–computed tomography (CT) using Fluoro Deoxy Glucose (FDG) [3]. These cancers are very hard to cure. But for the chance of survival, the early detection of this carcinoma or cancer is very necessary with the help of computer aided diagnosis system.

II. PROPOSED FLOW

The proposed method provides the automated diagnosis system for lung cancer with high detection rate and minimum computation time. Generally this method consists of four major steps ie, 1) Enhancement 2) Segmentation 3) Hybrid feature generation 4) Classification. The enhancement and hybrid feature generation leads to improve the detection result. Move over the computation time also reduced with respect to generated features compared to other existing diagnosis methods. The skewness based image enhancement will reveal the fine information to detect the cancer region pretty good. The segmentation helps to separate the lung area from the chest wall for further proceeding like feature generation and

---

[1] *Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India*
[2] *Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India*

classification Also the feature normalisation will be done to make all the features in a certain range also to improve he training status of the SVM classifier, and the hybrid feature set generation helps to improve the detection accuracy for better diagnosis.

A. ENHANCEMENT - CONTRAST LIMITED ADAPTIVE HISTOGRAM

While performing AHE, it is quite acceptable that noise in the image gets enhanced. It can bring about some sort artifacts to appear on those regions. To constrain the presence of such artifacts and noise, a variation of AHE called CLAHE can be utilized. The amount of contrast enhancement for some intensity is directly inline to the slope of the CDF function at that intensity. Hence contrast enhancement can be controlled by restricting the slope of the CDF. The slope at each bin is determined by its histogram height. In this way controlling the height limits the slope of the CDF and thereby contrast enhancement is achieved. The main contrast between normal AHE and CLAHE is that there is one additional step to clip the histogram before the calculation of its CDF while mapping function is performed. Following is the step by step explanation of the algorithm for this function:
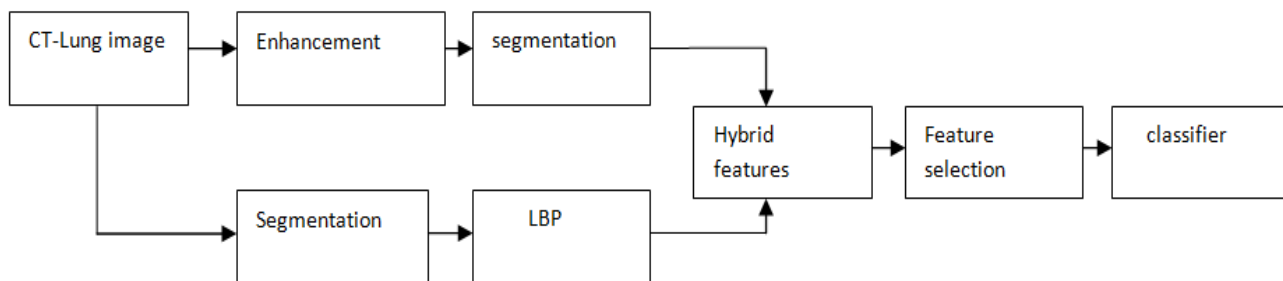

Figure 1. Proposed Architecture Based On Classifiers

So when we combine these two mathematical functions we can get the most suitable enhancement factor in between zero and one. It gives the change in the intensity of original pixels. Figure 2.2. shows the enhancement result of different normal and abnormal Lung CT images.
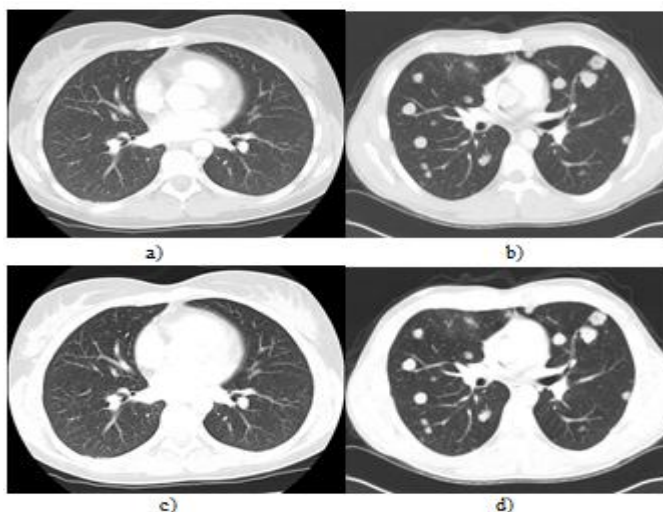

Figure 2. a) Normal, b) Abnormal, c) Normal Enhanced, d) Abnormal Enhanced image

B. SEGMENTATION

Adaptive thresholding followed by morphological operations is done to separate the lung region from chest wall, its mathematically represented by

$$F(x, y) = I(x, y) * T(x, y) \qquad (1)$$

$$G(x, y) = Id(x, y) * F^{''}(x, y) \qquad (2)$$

Where, T(x,y) represents the threshold, Id(x,y) represents dilation and $F^{''}$(x,y) denotes the complement of an image. Here the Adaptive thresholding [14] performed based on otsu thresholding, it separates the lung region from the background. Each pixel intensity have different threshold so that it makes the segmentation even better. Even after thresholding, because of change in intensities of inside lung region, some holes may be present. Inorder to fill that region, morphological operation is performed to fill those regions. The morphological dilation operation done by disc structuring element with the radius 3, so that exact lung region can be extracted from chest wall.

III. FEATURE GENERATION

The hybrid feature is the combination of specific set of feature descriptor which is taken from direct and enhanced images. These TOF (Topological Oriented Feature), POF (Pattern Oriented Feature), and FOG (Frequency Oriented Feature) ,are the specific set of feature descriptors which is used to train the classifier. These three set of features are discriminating the geometrical, spatial arrangements and coarse information of the images. Especially The TOF gives the geometrical oriented description and continuous change in shape. The POF gives the consistency of the surface or the spatial arrangement of the gray levels. Finally the FOG helps to generate the variation in the different subband analysis in the form of frequency. These set of features leads to improve the classifier performance.

a)   Topological Oriented Features

The Topological oriented features are Delaunay Triangulation and Voronoi Diagram. The Triangulation is a matrix representing the set of simplexes that make up the triangulation. It defines the constrained edge data in the triangulation. The matrix provides the number of simplexes and the amount of vertices per simplex. The triangulation is signified by standard simplex-vertex format; each row specifies a simplex defined by indices into X.

b)   Frequency Oriented Features

The wavelets used here generate the FOF features. The discrete wavelets allow analysing the temporal evolution of the frequency content of the image, The SWT (Stationary Wavelet Transform) captures both the spatial and frequency information of a signal. It analyses the image by decomposing it into a coarse estimate via low-pass filtering and into feature information via high-pass filtering. It is adequate that we consider four decomposition directions corresponding to $0\circ$ (*Dh*), $45\circ$ (*Dd*), $90\circ$ (*Dv*), and $135\circ$ (*Dd*) orientations.

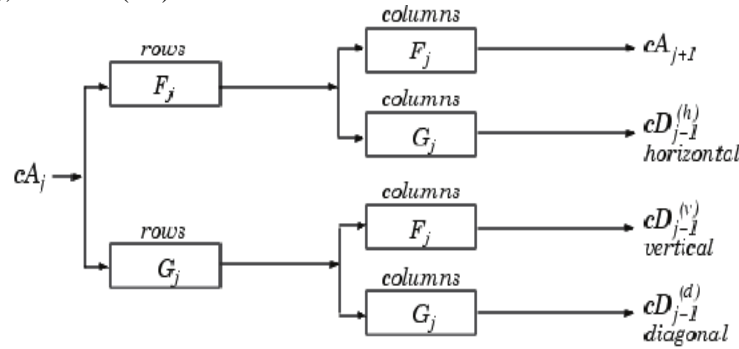*Figure 3. SWT Decomposition*

The four coefficients are computed by the following

$$A_{j,k_1,k_2} = \sum_{n1}\sum_{n2} h_0 \uparrow^{2j}(n_1-2k_1)h_0 \uparrow^{2j}(n_2-2k_2)A_{j-1,n_1,n_2} \qquad (3)$$

$$H_{j,k_1,k_2} = \sum_{n1}\sum_{n2} h_0 \uparrow^{2j}(n_1-2k_1)g_0 \uparrow^{2j}(n_2-2k_2)A_{j-1,n_1,n_2} \qquad (4)$$

$$V_{j,k_1,k_2} = \sum_{n1}\sum_{n2} h_0 \uparrow^{2j}(n_1-2k_1)h_0 \uparrow^{2j}(n_2-2k_2)A_{j-1,n_1,n_2} \qquad (5)$$

$$D_{j,k_1,k_2} = \sum_{n1}\sum_{n2} h_0 \uparrow^{2j}(n_1-2k_1)g_0 \uparrow^{2j}(n_2-2k_2)A_{j-1,n_1,n_2} \qquad (6)$$

where, where, $g_0 \uparrow^{2j}$ represents the low pass filter, $h_0 \uparrow^{2j}$ represents the high pass filter, S refers to the input image and (n_1-2k_1) is the scaling factor. The features calculated for different sub bands is given below

$$Energy = \frac{1}{p^2 \times q^2}\sum_{x=\{p\}}\sum_{y=\{q\}}|Dv_1(x,y)|^2 \qquad (7)$$

$$Entropy = -\sum_{x=\{p\}}P(Dv_1)\log(Dv_1) \qquad (8)$$

$$Average = \frac{1}{p \times q}\sum_{x=\{p\}}\sum_{y=\{q\}}|Dv_1(x,y)| \qquad (9)$$

IV. FATURE NORMALISATION

These all set of features is subject to *z*-score normalization. In the process of *z*-score normalization, the feature vector is converted to zero mean and unit variance. The mean and standard deviation of the input vector are computed as follows:

$$F_{new} = \frac{F_{old}-mean}{std} \qquad (10)$$

Where, *Fold* is the original value, *F*new is the new value and the *mean* and *std* are the mean and standard deviation of the original data range respectively. Figure. 6 shows the *z*-scored normalized distribution for each of the features across the 100 cancerous and 120 normal samples used in this study.
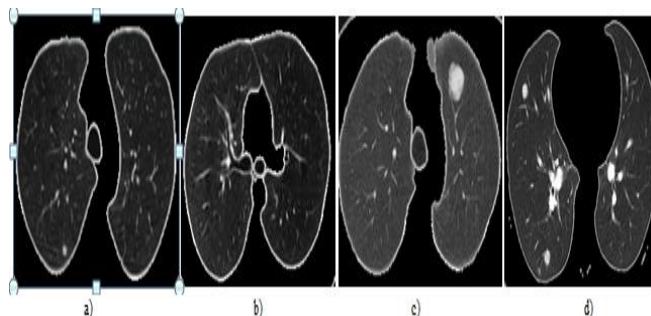

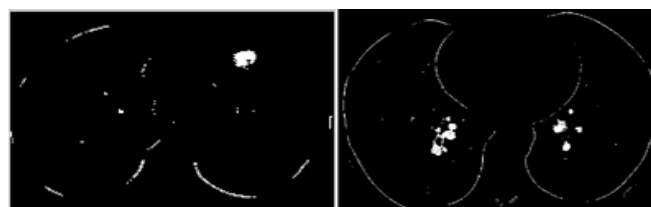Figure 4.   Segementation of a-b) Normal c-d) Abnormal Enhanced lung region

Figure 5. Suspecious regions of two lung images

V. CLASSIFICATION

The selection of classifier is very important for accurate classification. Both the supervised and unsupervised classifiers are used for classification but the supervised classifier outperforms the other. Here we use SVM [13], [6] classier with RBF kernel to classify the input feature data. Tables show that the classifier detection rate and sensitivity and specificity of SVM are very good compared to other classifiers.

VI. EXPERIMENT AND RESULT

We have done our proposed method for 40 different patients with the training data of 120 normal and 100 abnormal samples. The detection rate exceeds 94 percentage, which shows the optimal rate to diagnose the cancerous CT lung images. Table 1 shows the enhancement achievement like PSNR, MSE and IEF. It achieves 10 for the lung images

Table.1 Performance evaluation for enhancement

Method	PSNR	MSE	IEF
Histogram	38	0.5	0.6
Gaussian	40	0.3	0.7
Hessian	42	0.35	0.75
A,Histogram	45	0.2	0.8
Proposed	47.8	0.15	0.9

Table. 2 Comparison among classifiers

Different classifier	Accuracy	Sensitivity
KNN	83.5	78
PNN	85.5	82
BAYES	87	85
FF	88	87
SVM	93	90

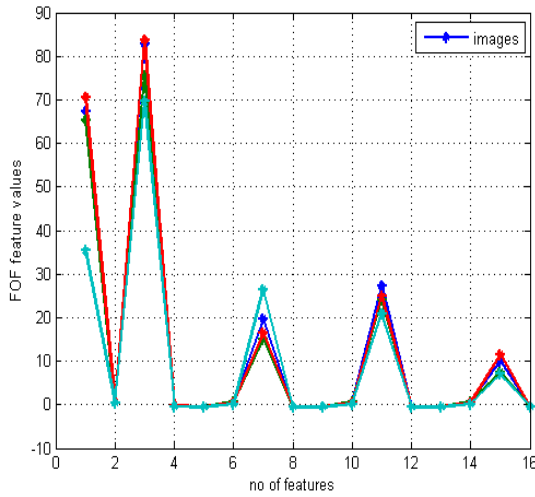Table.6.3 Comparison among the existing methods
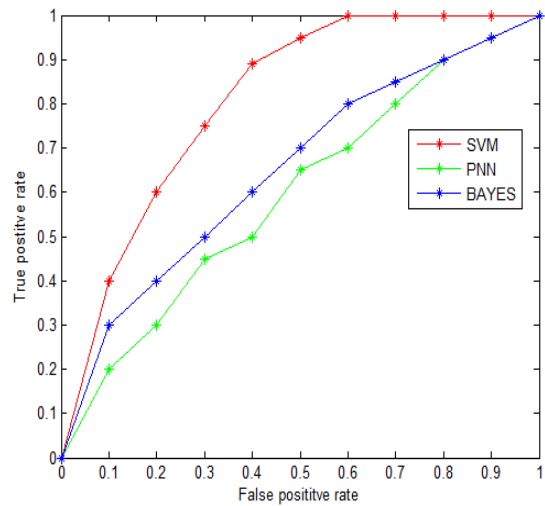
Figure 6. Eight FOF features among various images

Figure 7. True positive vs false positive rate
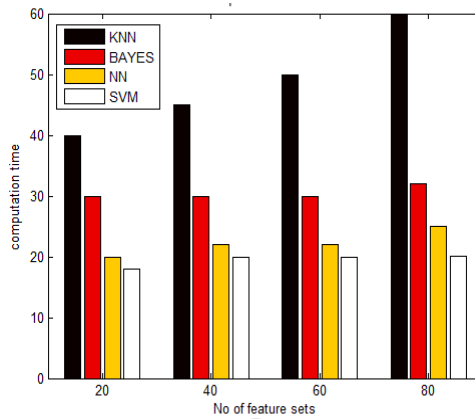

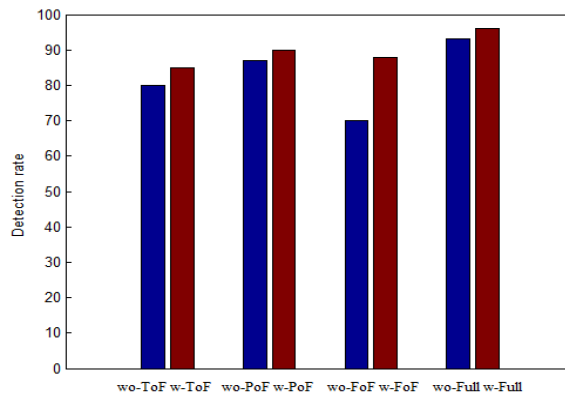*Figure 8. Comparison of computational time when finding the number of feature sets among classifiers*

Figure 9. Comparison of detection rates (wo-ToF--with out Topological Features ,w-ToF--with Topological Features)

VII. CONCLUSION

In this paper, we propose a new method for finding the lung cancer in CT images with the help of SVM classifier and hybrid features. The proposed hybrid features helps to improve the classification accuracy so this method is well suitable for detecting the lung cancer from the given medical image. Also this method performs well compared to other conventional methods with respect to detection rate and computation time.

REFERENCES

[1] Samuel h. Hawkins, john n. Korecki1, yoganand balagurunathan2, yuhua gu2, Virendra kumar2, satrajit basu1, lawrence(2014), predicting outcomes of nonsmall cell lung Cancer using ct image features, IEEE access 2014

[2] Ganeshan, B., Abaleke, S., Young, R. C., Chatwin, C. R., & Miles, K. A. (2010). Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer imaging, 10(1), 137-143.

[3] Samala, R., Moreno, W., You, Y., & Qian, W. (2009). A novel approach to nodule feature optimization on thin section thoracic ct. Academic radiology, 16(4), 418-427.

[4] Way, T. W., Hadjiiski, L. M., Sahiner, B., Chan, H. P., Cascade, P. N., Kazerooni, E. A.,& Zhou, C. (2006). Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. Medical Physics, 33(7), 2323-2337.

[5]   Lee, M. C., Boroczky, L., Sungur-Stasik, K., Cann, A. D., Borczuk, A. C., Kawut, S. M., & Powell, C. A. (2010). Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction. Artificial intelligence in medicine, 50(1), 43-53.

[6]   Zhu, Y., Tan, Y., Hua, Y., Wang, M., Zhang, G., & Zhang, J. (2010). Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography. Journal of digital imaging, 23(1), 51-65.

[7]   Al-Kadi, O. S., & Watson, D. (2008). Texture analysis of aggressive and nonaggressive lung tumor CE CT images. IEEE transactions on biomedical engineering, 55(7), 1822-1830.

[8]   Kido, S., Kuriyama, K., Higashiyama, M., Kasugai, T., & Kuroda, C. (2003). Fractal analysis of internal and peripheral textures of small peripheral bronchogenic carcinomas in thin-section computed tomography: comparison of bronchioloalveolar cell carcinomas with nonbronchioloalveolar cell carcinomas. Journal of computer assisted tomography, 27(1), 56-61.

[9]   Segal, E., Sirlin, C. B., Ooi, C., Adler, A. S., Gollub, J., Chen, X. & Kuo, M. D. (2007). Decoding global gene expression programs in liver cancer by noninvasive imaging. Nature biotechnology, 25(6), 675-680.

[10]   De_niens Developer XD 2.0.4 User Guide, Deniens AG, München, Germany, 2009.

[11]   Gu, Y., Kumar, V., Hall, L. O., Goldgof, D. B., Li, C. Y., Korn, R., &Lambin, P. (2013). Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. Pattern Recognition, 46(3), 692-702.

[12]   H. W. Tsao and Y. C. Fan, "Method and device for IC identification", R.O.C. Patent I226 001, Jan. 1, 2005.

[13]   Arumuga Maria Devi, Mebin Jose V. I, Kumar Parasuraman P, A novel approach for automatic Non small cell lung carcinoma in CT images. International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), (2016).

[14]   Arumuga Maria Devi , Implementation of Lung Cancer Nodule Feature Extraction using Threshold Technique, International Advanced Research Journal in Science, Engineering and Technology, Vol. 3, Issue 8, August 2016

[15]   Arumuga Maria Devi, Hyperspectral Image Classification using Spatial and Spectral Features, International Journal of Scientific and Engineering Research, Jul 2013

[16]   Arumuga Maria Devi ,Graph Cut Based Method for Automatic Lung Segmentation for Tuberculosis by using Screening Method in Chest Radiographs,Ciit International journal of digital image processing Volume 7, No 09, October – November 2015

[17]   Arumuga Maria Devi ,A Novel Technique of Resolution Enhancement in Hyperspectral Images on Proposed CHLAE Technique, Journal of Chemical and Pharmaceutical Sciences (JCPS) Scopus indexed journal Volume 9, Issue 1, January-March 2016

[18]   Arumuga Maria Devi ,Automatic Liver Segmentation using Mean Shift Techniques, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)   Volume 3 Issue 11, November 2014

Dr. T. Arumuga Maria Devi Received B.E. Degree in Electronics & Communication Engineering from Manonmaniam Sundaranar University, Tirunelveli India in 2003, M.Tech degree in Computer & Information Technology from Manonmaniam Sundaranar University, Tirunelveli, India in 2005 and Worked as Lecturer in department of Electronics & Communication Engineering in Sardar Raja College of Engineering and also received Ph.D Degree in Information Technology – Computer Science and Engineering from Manonmaniam Sundaranar University, Tirunelveli , India in 2012 and also the Assistant Professor of Centre for Information Technology and Engineering of Manonmaniam Sundaranar University since November 2005 onwards. Her research interests include Signal and Image Processing, Multimedia and Remote Communication.

Mebin Jose V.I was born in Nagercoil, india, R.O.C., in 1988. He received the B.E.degree in Electronics and Communication Engineering from CSI Institute of technology Thovalai in 2009 and done his M.tech in 2011 at Karunya University ,Coimbatore , currently he is doing Phd in Centre for Information Technology and Engineering , MS University, India,, His research interests are Digital and Medical image processing,multi media communication, machine vision ,machine learning,wireless communication.