Computerised Tuberculosis Detection Using Artificial Neural Network

Saandthra Anna Jose[1], Kavitha N. Nair[2]

[1, 2]Mahatma Gandhi University, University College of Engineering, Thodupuzha, Kerala 685 587, India

Abstract: *Tuberculosis (TB) is an infectious disease caused by the bacillus Mycobacterium tuberculosis, which typically affects the lungs. Diagnosing TB is still a major challenge. The definitive test for TB is the identification of Mycobacterium tuberculosis in a clinical sputum or pus sample, which is the current gold standard. However, it may take several months to identify this slow-growing organism in the laboratory. Another techniques are sputum smear microscopy, skin tests based on immune response, but they are not always reliable. In an effort to reduce the burden of the disease, develop a computerised tuberculosis detection using artificial neural network. In this thesis, a new method for detecting normal and abnormal TB CXR by using Artificial Neural Network (ANN) classifier is presented to improve the accuracy and performance of the system. An automated system that detect the TB from a set of CXR, the basic procedure behind this system is training and testing. Firstly extract the lung region using an optimization method based on graph min cut, then find the features of the lung region of all normal and abnormal CXR. For each normal and abnormal CXR have object oriented inspired features and CBIR based features. The object oriented inspired features include intensity histogram, gradient magnitude histogram, HOG and LBP. The CBIR based features are colour,edge,fuzzy histogram, Hu moments and autocorrelation. Each feature descriptor is quantized into 8 bin histograms overall number of features is thus 9*8=72.So each input image has a 72 bin feature vector. This 72 bin feature vector of all CXR is one of the input of the ANN classifier for training purpose. During testing extract the lung region and find the all feature vector of the input test CXR ,this feature vector is the second set of input to the ANN classifier, which enables the region of all normal and abnormal CXR. For each normal and abnormal CXR have object oriented inspired features and CBIR based features. The object oriented inspired features include intensity histogram, gradient magnitude histogram, HOG and LBP. The CBIR based features are colour,edge,fuzzy histogram, Hu moments and autocorrelation. Each feature descriptor is quantized into 8 bin histograms overall number of features is thus 9*8=72.So each input image has a 72 bin feature vector. This 72 bin feature vector of all CXR is one of the input of the ANN classifier for training purpose. During testing extract the lung region and find the all feature vector of the input test CXR ,this feature vector is the second set of input to the ANN classifier, which enables the CXR to be classified normal or abnormal using artificial neural network.*

Keywords: Graph cut segmentation, object based oriented features, content based image retrieval features, artificial neural network classification

1. Introduction

Various methods used for identification of tuberculosis, currently using methods are identification of mycobacterium tuberculosis in a clinical sputum or pus sample, skin test through blood test and manual analysis of chest X-ray(CXR) and Computer Tomography(CT) scan. They are slow and often unreliable. In the existing system, automated approach for detecting tuberculosis in conventional poster anterior chest radiographs. First extract the lung region using segmentation method. For this lung region compute a set of texture and shape features, which enable the X-rays to be classified as normal or abnormal using a binary classifier [1]. Graph cut segmentation is used for extracting lung region. Lung segmentation is the process of extracting lung region from the input CXR using edge detection in combination with graph cut method. First find out the edges of the lung region. An Edge in an image is a significant local change in the image intensity, usually associated with a discontinuity in either the image intensity or the first derivative of the image intensity.

To describe normal and abnormal patterns in the segmented lung field, there are two different feature sets one is object detection inspired features-set A and other is CBIR based image features.[1].

Artificial neural network used as a classifier. Artificial neural networks are relatively crude electronic networks of "neurons" based on the neural structure of the brain. They process records one at a time, and "learn" by comparing their classification of the record (which, at the outset, is largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm the second time around, and so on for many iterations.

The rest of this paper organized as follows. Section II reviews the related works in computerised TB and other lung disorders. The TB detection using artificial neural network is explained in section III. Then experimental results are shown in section IV. Finally, conclusions are presented in section V.

2. Related Work

Various segmentation methods and classifiers is used for Automatic Tuberculosis Screening Using Chest Radiographs Automated approach for detecting TB manifestations in chest X-rays (CXRs), based on lung segmentation and lung disease classification. Lung segmentation ,feature extraction and classification are the main methods used . lung segmentation as an optimization problem that takes properties of lung boundaries, regions, and shapes into account. First train the SVM classifier using number of

normal and abnormal CXR, training includes both lung segmentation and feature extraction .Features include Object Detection Inspired Features and content-based image retrieval features. Next step is the testing of the input chest radiographs. In testing, extract the lung region of input CXR using graph cut segmentation method in combination with lung model. From the segmented lung model ,system derive the set of textures and shape features as input to the pre trained SVM classifier. To detect abnormal CXRs with TB, SVM which classifies the computed feature vectors into either normal or abnormal using decision rules and thresholds. An SVM in its original form is a supervised non probabilistic classifier that generates hyper planes to separate samples from two different classes in a space with possibly infinite dimension[1].

The other main approaches for lung segmentation are rule-based reasoning and pixel classification with neural networks or other classifiers[2]. In graph cut segmentation, the lung mask constructed by removing the airway mask from the segmented lungs can be a single connected component, for instance when lungs are closed in some regions. To find where to cut our mask, we used a graph-cut algorithm that separates two seed regions on the optimal border over a certain neighbourhood constraint. The graph-cut implementation, which finds on a weighted graph which edges to cut in order to separate the source seeds and the sink seeds. The input graph is composed of all voxels of the connected lung label, which is constructed by repeatedly closing the lung mask until having one single large component (which, in practice, represents more than 99% of the size of the union of all components in the image). Each of these voxels becomes a node of the graph and is connected to its neighbours by edges labelled with the inverse of the gradient of the initial image. Source and sink nodes are obtained by eroding the closed mask until having two distinct regions, which are labelled as left and right seeds. The graph-cut result cuts a closed mask which can be larger than the lungs in some places. The exceeding voxels are removed by masking the graph-cut result with the lung mask. Both lung masks are summed to obtain the final result

Neural network classifier, classification is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. Neural networks are data driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximators in that neural networks can approximate any function with arbitrary accuracy. Since any classification procedure seeks a functional relationship between the group membership and the attributes of the object, accurate identification of this underlying function is doubtlessly important. Third, neural networks are nonlinear models, which makes them flexible in modelling real world complex relationships. Finally, neural networks are able to estimate the posterior probabilities, which provides the basis for establishing classification rule and performing statistical analysis. On the other hand, the effectiveness of neural

network classification has been tested empirically. Neural networks have been successfully applied to a variety of real world classification tasks in industry, business and science. In addition, several computer experimental evaluations of neural networks for classification problems have been conducted under a variety of conditions.

3. Proposed Method

The main procedures in the proposed system are collection large set of normal and abnormal CXR ,segmenting the lung region of CXR using graph cut lung segmentation. Then extract the features from segmented lung region and features is used to train the artificial neural network classifier. When apply the test images, the classifier compares the features of the test images with trained image features and make the decision as test image is normal or abnormal.
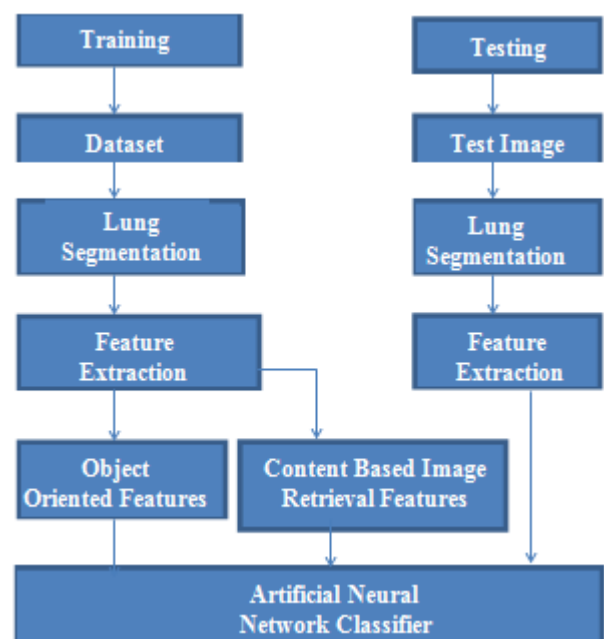

Figure 1: Block diagram of proposed model

3.1. Lung Segmentation Using Graph Cut

Lung segmentation is the process of extracting lung region from the input CXR using edge detection in combination with graph cut method. First find out the edges of the lung region. An Edge in an image is a significant local change in the image intensity, usually associated with a discontinuity in either the image intensity or the first derivative of the image intensity. Many points in an image have a nonzero value for the gradient, and not all of these points are edges for a particular application. Thresholding is used for the detection of edge points. The simplest thresholding methods replace each pixel in an image with a black pixel if the image intensity $I\{i,j\}$ is less than some fixed constant T (that is, $I\{i,j\}<T$), or a white pixel if the image intensity is greater than that constant. For calculating the graph cut, max-flow algorithm is widely used. Segmentation procedures includes,

- Specify seed pixels (on objects or background).
- A weighted directed graph among the pixels in the image is generated based on the seed pixels.

Volume 4 Issue 12, December 2015

Paper ID: NOV152218   1181

- A min-cut of the weighted directed graph is calculated using max-flow algorithm.

Graph Cut Algorithm
V is a set of vertices (pixels), and it includes two special nodes, s and t. E is a set of edges between two vertices, and each edge has a non-negative capacity $c(u,v)$, Then,

- G=(V,E)   (1)
  a weighted directed graph
- A cut of the graph is the division of V into two groups, S and T.
- The capacity of the cut c(S,T) is defined as the capacity of the edges from S to T.
- The cut which minimizes c(S,T) is called min-cut.
- The weighted directed graph is generated from the pixels in the image.
- Making a seed: a black pixel is specified as foreground, and a white pixel is specified as background.
- Then, pixels that have similar colour to black have strong connection to s. On the other hand, pixels that have similar colour to white have strong connection to t.
- Making a graph(Energy): Min-cut corresponds to the minimum energy.
- Compute Min-cut-according to ―max-flow min-cut theorem", min-cut is obtained from the result of max-flow. In order to calculate max-flow, ―augmenting path method" scans the graph to find a path from source (s) to sink (t).

3.2. Feature Extraction

To describe normal and abnormal patterns in the segmented lung field, there are two different feature sets.one is object detection inspired features-set A and other is CBIR based image features [1].

Object Detection Inspired Features-Set A: First set, we use features that we have successfully applied to microscopy images of cells for which classified the cell cycle phase based on appearance patterns. The first set is a combination of shape, edge, and texture descriptors. For each descriptor, compute a histogram that shows the distribution of the different descriptor values across the lung field. Each histogram bin is a feature, and all features of all descriptors put together form a feature vector that is the input to the classifier. Through empirical experiments, fusing 32 bins for each histogram gives us good practical results, shape and texture descriptors as following

- Intensity histograms (IH).
- Gradient magnitude histograms (GM).
- Histogram of oriented gradients (HOG).

HOG is a descriptor for gradient orientations weighted according to gradient magnitude.

- Local binary patterns (LBP)

LBP is a texture descriptor that codes the intensity differences between neighboring pixels by a histogram of binary patterns. LBP is thus a histogram method in itself. The binary patterns are generated by thresholding the relative intensity between the central pixel and its neighbouring

pixels. Because of its computational simplicity and efficiency, LBP is successfully used in various computer vision applications, often in combination with HOG [1].

CBIR Based Image Features –Set B : The second feature set, Set B is a group of low-level features motivated by content-based image retrieval (CBIR) .Feature Set B contains the following

- Colour and edge direction descriptor (CEDD)

CEDD is a low level feature that is extracted from the images and can be used for indexing and retrieval. This feature is called ―Colour and Edge Directivity Descriptor" and incorporates colour and texture information in a histogram. CEDD size is limited to 54 bytes per image, rendering this descriptor suitable for use in large image databases. One of the most important attribute of the CEDD is the low computational power needed for its extraction.

- Fuzzy colour and texture histogram (FCTH)

CEDD and FCTH give the colour and texture information in one histogram. They differ in the way they capture texture information.

- Hu moments

These moments are widely used in image analysis. They are invariant under image scaling, translation, and rotation. The distributed content-based visual information retrieval(DISCOVIR) system is used to extract Hu moments.

- Edge and autocorrelation

These are well-known texture analysis methods, which use statistical rules to describe the spatial distribution and relation of gray values [1]. Texture can be defined as a regular repetition of an element or pattern on a surface. Image textures are complex visual patterns composed of entities or regions with sub-patterns with the characteristics of brightness, color, shape, size, etc. An image region has a constant texture if a set of its characteristics are constant, slowly changing or approximately periodic[4].

3.3. Classification

Artificial neural network used as a classifier. Artificial neural networks are relatively crude electronic networks of "neurons" based on the neural structure of the brain. They process records one at a time, and "learn" by comparing their classification of the record (which, at the outset, is largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm the second time around, and so on for many iterations.

4. Simulation results and discussion

Chest radiographs are of paramount importance in the identification of patients with abnormal pulmonary conditions [2].The TB database consist of two sets of CXR. One set is used for training the classifier and the other set for testing purpose. TB first database contains abnormal and

normal posterior– anterior (PA)CXR collected from various hospitals. In order to be able to train classifiers to distinguish between normal and abnormal images, radiologists have outlined the areas containing abnormalities in each abnormal image. During this procedure, several abnormal images were re-classified by the radiologist as normal. The second database contains normal and abnormal images for testing purpose.

The lung segmentation is the process of extracting the lung region from the CXR using edges of the image and energy associated with randomly selected seeds. The edges of the image can be finding out by normal thresholding. In normal thresholding methods replace each pixel in an image with a black pixel if the image intensity I-$\{i,j\}$ is less than some fixed constant T (that is, $I_{\{i,j\}}<T$), or a white pixel if the image intensity is greater than that constant. The lung segmentation can be carried out in the trained input images as well as test images. The Fig. 2 showing the normal input image and Fig. 3 is the corresponding graph cut segmented lung region. The Fig.4 is the abnormal CXR and Fig.5 represents the corresponding graph cut segmented lung region.
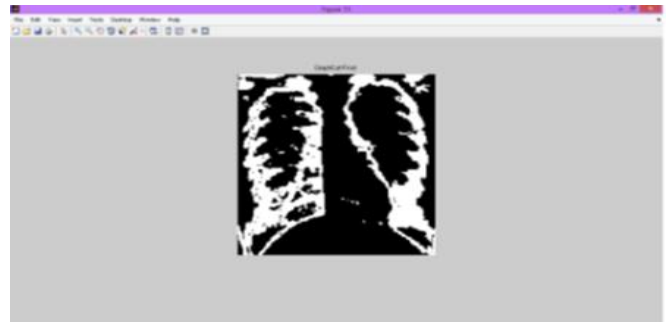

Figure 2: Normal CXR image

Figure 3: Graph cut segmented lung region of normal CXR

Figure 4: Abnormal CXR

Figure 5: Graph cut segmented lung region of abnormal CXR

Feature Extraction: For each normal and abnormal CXR extract the object oriented inspired features and CBIR based features. The object oriented inspired features include intensity histogram, gradient magnitude histogram, HOG and LBP.The CBIR based features are color,edge,fuzzy histogram, Hu moments and autocorrelation. Each feature descriptor is quantized into 8 bin histograms overall number of features is thus 9*8=72.So each input image has a 72 bin feature vector. The feature vector bins 1 to 8 represents intensity histogram,9 to 16 is the gradient magnitude histogram,17 to 24 HOG features,25 to 35 LBP,33 to 40 color features,41 to 48 edge features,49 to 56 fuzzy color features, 57 to 64 Hu moments,65 to 72 autocorrelation. Fig. 6 and Fig. 7 showing the 1*72 double feature vector of a normal and abnormal CXR image, and Figure 4.7 represents the feature vector of all normal and abnormal CXRs used for training.

Figure 6: Feature vector of a normal image

Figure 7: Feature vector of an abnormal Image

Paper ID: NOV152218

1183

Figure 8: Feature vectors for all normal and abnormal CXR for training

Training: Artificial neural network is trained with both normal and abnormal CXR database. Before training, lung region of all the CXR data base is segmented and extract the features. The normal and abnormal CXR have different range of feature vectors. The artificial neural network is trained with various normal and abnormal feature vectors. The artificial neural network is trained with 14 normal and abnormal CXR data base. The Fig. 9 represents the neural network is trained with feature vectors of all normal and abnormal CXR
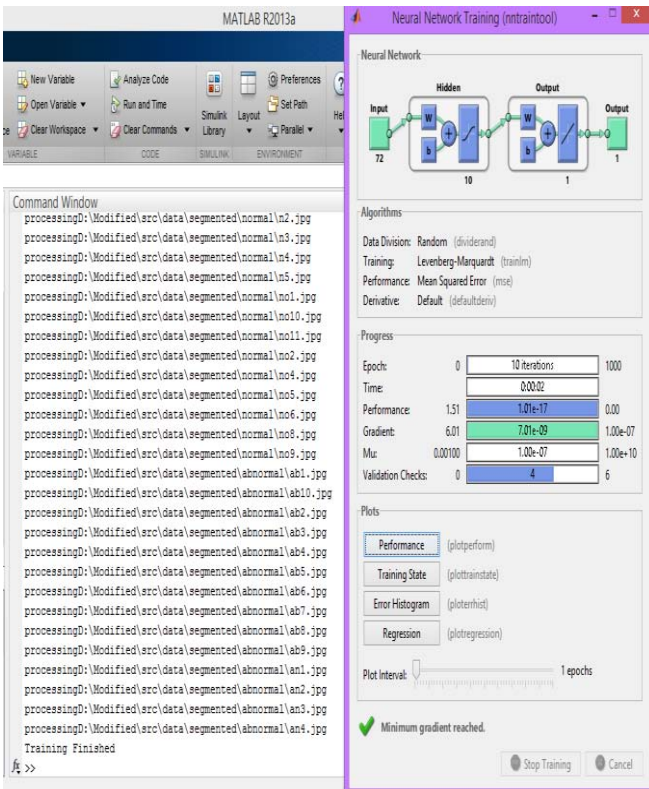

Figure 9: Artificial Neural Network Training

Testing: The artificial neural network is used as a classifier, which compares the features are extracted from the

segmented lung region of the test image and the trained feature vector values. Finally the artificial neural network which classifies the computed feature vectors of the test image is either normal or abnormal. Fig.10 showing the normal test image and Fig.11 is the representation of abnormal test image CXR on command window.
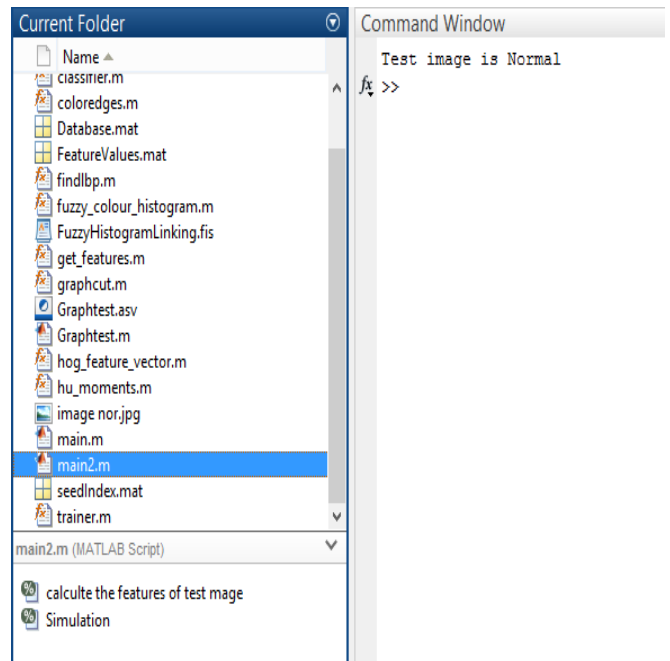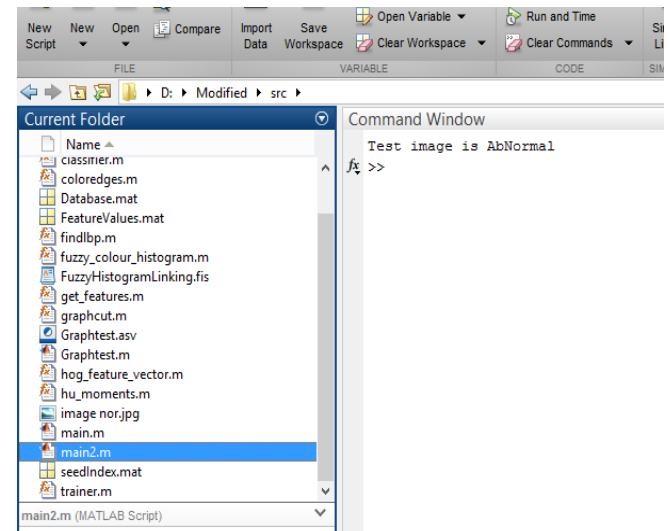

Figure 10: Command window of normal CXR image

Figure 11: Command window of abnormal CXR image

5. Conclusion

In this paper, a new method for detecting normal and abnormal TB CXR by using Artificial Neural Network (ANN) classifier to improve the accuracy and performance of the system. A computerised system that detect the TB from a set of CXR, basic procedure behind this system is training and testing. Firstly extract the lung region using an optimization method based on graph min cut algorithm. Then find the features of the lung region of all normal and abnormal CXR. For each normal and abnormal CXR have object oriented inspired features as first feature set and CBIR based features as second set. The object oriented inspired

Paper ID: NOV152218   1184

features include intensity histogram, gradient magnitude histogram, HOG and LBP. The CBIR based features are color, edge, fuzzy histogram ,Hu moments and autocorrelation. The second feature set, is a group of low-level features motivated by content-based image retrieval (CBIR). This feature collection includes intensity, edge, texture and shape moment features, which are typically used by CBIR systems. The entire feature vector has 40 dimensions, which is more than the feature vector of Set A, and which allows us to evaluate the effect of high-dimensional feature spaces on classification accuracy. Each feature descriptor is quantized into 8 bin histograms overall number of features is thus 9*8=72.So each input image has a 72 bin feature vector. This 72 bin feature vector of all CXR is one of the input of the ANN classifier for training purpose. \During testing extract the lung region and find the all feature vector of the input test CXR, this feature vector is the second set of input to the ANN classifier. The classifier has two sets of input data, one is the extracted object detection inspired features and CBIR based image features of the trained database which is stored as feature vector in the classifier for further application. The other one is the extracted features of the test input. The artificial neural network compares the values of feature vectors in those two set using some decision rules and thresholds, which classifies the computed test image feature vectors into either normal or abnormal.

To improve the performance further, try to improve the lung segmentation, which provides average performance. One approach would be to find optimal weights for the terms in the graph cut energy function. Another possibility would be to use atlas-based lung models for computing the average lung model, but it is time consuming and increases the complexity of the system.

References

[1] Stefan Jaeger, Alexandros Karargyris, Sema Candemir,Les Folio,"Automatic tuberculosis screening using chest radiographs", IEEE Transactions on Medical imaging, vol. 33 , no. 2, pp. 233-245, February 2014.

[2] Bram van Ginnekena and Bart M. ter Haar Romeny, "Automatic segmentation of lung fields in chest radiographs, ―Med. Phys., vol. 27, no. 10, pp.2445–2455, 2000.

[3] Reza Devarzani,Saeed Mozaffari and Khashayar Yaghmaie," Scale- and rotation-invariant texture description with improved local binary pattern features",Signal Processing,vol. 111, pp. 274-293, June 2015.

[4] T.Ojala,M.Pietikäinen,andT.Mäenpää, ―Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp.971-987, Jul. 2002.

[5] G. Srinivasan and G. Shobha, ―Statistical texture analysis," Proc. World Acad. Sci., Eng. Technol., vol. 36, pp. 1264–1269, 2008.

[6] S. Chatzichristofis and Y. Boutalis, ―FCTH: Fuzzy color and texture histogram—A low level feature for accurate image retrieval," in Proc. Int. Workshop Image Anal. Multimedia Interactive Services, 2008, pp. 191–196.

Author Profile

Saandthra Anna Jose, B.Tech degree in Applied Electronics & Instrumentation Engineering from Adi Shankara Institute of Engineering &Technology, Kalady, India in the year 2012. At present she is pursuing M.Tech. with the specialization of Applied Electronics in Mahatma Gandhi University College of Engineering, Thodupuzha, India.

Kavitha.N.Nair completed her M.Tech in Applied Electronics from Mahatma Gandhi University College of Engineering, Thodupuzha in 2012. At present she is working as lecturer in Electronics in Department of Electronics and Communication Engineering, Mahatma Gandhi University College of Engineering, Thodupuzha, India.