COMPUTER AIDED DETECTION OF TUBERCULOSIS USING AUTOMATIC LUNG SEGMENTATION AND HYBRID CLASSIFIER

Sethu Parvathi C[1], Chithra A S[2]

[1]M- Tech Student, [2]Associate Professor

Abstract

Tuberculosis is a serious health threatening disease which is more prevalent in African countries. According to WHO, it is the second leading cause of death from an infection after HIV and based on the estimates of Centre for Disease Control and Prevention (CDC), 10.4 million cases of TB and 1.7 million deaths were reported worldwide in 2016. Earlier TB detection methods do not ensure accurate results and hence need proper tools to detect the disease. To address this issue, a novel mechanism is proposed to automatically detect tuberculosis in Chest X-Rays. The proposed system comprises of a three stage detection process. The first stage is the segmentation using adaptive thresholding followed by EM segmentation. The second stage is feature extraction which extracts set of features from an input image and the final stage comprises of a trained classifier which detects normal or abnormal images. The classifier used in this system is a hybrid of k-Nearest Neighbor Classifier (k- NN) and Support Vector Machine (SVM). Performance analysis is based on four performance analyzing criteria- accuracy, sensitivity, specificity, area under ROC curve (AUC). A comparison task is also performed to measure the quality of classifiers among two classifiers – k-NN, hybrid classifiers.

Index Terms: Tuberculosis, Computer- aided Detection, Lung Segmentation, Feature extraction, Classification.

I. INTRODUCTION

Tuberculosis is a potentially serious infectious disease caused due to the inhalation of an infectious bacteria- *Mycobacterium Tuberculosis.* According to the estimates of Centre for Disease Control and Prevention (CDC), 10.4 million cases of TB and 1.7 million deaths were reported worldwide in 2016. Surprisingly, it is the second biggest killer from an infectious disease after HIV. This disease that usually affects the lungs is spread through air from person to person.

There are two kinds of tuberculosis infection- latent TB and active TB. In latent TB, the bacteria remain in the body in an inactive state. They cause no symptoms and are not contagious, but they can become active. In active TB, the bacterium causes symptoms and can spread from person to person. It may lead to death if not treated. About one-third of the world's population is believed to have latent TB. There is a 10 percent chance of latent TB becoming active, but this risk is much higher in people who have compromised immune systems i.e., people living with HIV or malnutrition, or people who smoke. The symptoms of active TB include tiredness, night sweats, no appetite, cough that lasts more than 3 weeks, fever, weight loss etc.

To check for TB, a doctor will use a stethoscope to listen to the lungs and check for swelling in the lymph nodes. The most common diagnostic test for TB is a skin test where a small injection of PPD tuberculin, an extract of the TB bacterium, is made just below the inside forearm. The injection site should be checked after 2-3 days, and, if a hard, red bump has swollen up to a specific size, then it is likely that TB is present. Unfortunately, the skin test is not 100 percent accurate and has been known to give incorrect positive and negative

readings. In sputum tests, cultured sputum samples are microscopically analyzed. Its low sensitivity and low specificity makes the method unreliable. Interferon Gamma Release Assays (IGRA) which is blood testing methods is of high cost and is rarely used. Out of several methods in detecting tuberculosis, chest X-Rays is found to be more reliable. Any manifestations of TB are clearly outlined and are relatively cheap. A chest x-ray (CXR) affected with TB will have cavitations, infiltrations, effusions or military patterns in the x-ray images. Examples of normal CXR and abnormal CXR (with TB infection) are shown in fig 1.1. (a) and fig 1.1. (b) respectively. The only solution to cure TB is proper diagnosis and treatment at the right time.
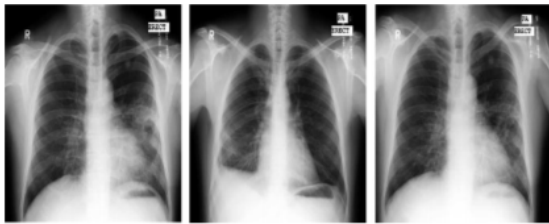

Fig 1.1 (a) Normal CXR

Fig 1.1 (b) Abnormal X-Ray -CXR - A is having a cavitary infiltrate in the left side and a subtle infiltrate in the right lower lung. CXR- B is having pleural TB. CXR- C is having infiltrates in both left and right lungs.

The proposed system focuses on developing an automatic TB detection system which accurately detects TB from chest radiographs. It can be explained through a three stage detection process.

1. Lung Segmentation
2. Feature Extraction
3. Trained classifier detects normal or abnormal images

Precisely, the proposed system is more reliable compared to other methods and is a simple way of detecting TB.

II. GENERAL FRAMEWORK FOR AUTOMATIC LUNG TUBERCULOSIS DETECTION SYSTEM

The general framework for automatic detection of tuberculosis consists of the following stages

1. Image Preprocessing
2. Image Segmentation
3. Feature Extraction
4. Classification
5. Decision Support

The first step in detection is Image Pre-processing. It is mainly done to enhance the image data and also to suppress unwanted distortions. Pre-processing also includes noise reduction, contrast enhancement, correction of pixels etc. The challenges possessed during TB detection are caused due to varying lung shape, strong edges of the rib cage, visible shape of the heart etc.

The second step is Image Segmentation which is done to extract good quality features. Two important properties are considered during segmentation – discontinuity and similarity. Discontinuity based segmentation is based on sharp intensity value changes. Similarity based segmentation on regions that are alike.

After segmentation, feature extraction is done on the segmented regions. The features that are extracted are used for identifying a normal or abnormal X- Ray image. The features are fed into the classifier which has got a trained set of normal and abnormal X-Ray images. Classifiers conclude whether the input image is normal or abnormal and the output is shown to the user. Figure 2 shows the flowchart representation of general framework of automatic tuberculosis detection system.
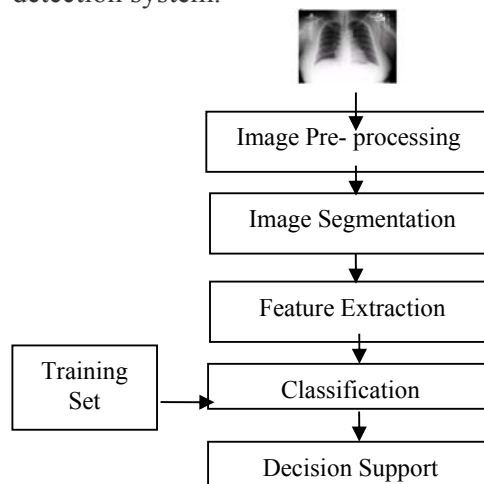

Fig 2. Flowchart representation of automatic tuberculosis detection system

III. RELATED WORK

The first step in lung detection is image pre-processing. The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image features important for further processing. Several reasons are mentioned for placing pre processing as a requirement for lung detection:

- To meet the requirements of physician for the image quality improvement
- Noise reduction
- Contrast enhancement
- Correction of missing or wrong pixel values
- Optimal preparation of data for post-processing
- Elimination of acquisition-specific artifacts

Processing of x-ray chest images poses some challenges. Figure 3.1 shows the variations of lung appearance in two different lung images. The challenges possessed during processing of chest X-Ray are varying lung shape, strong edges of the rib cage, visible shape of the heart, intensity variation around the clavicle bones and sharp corner at costophrenic angle.
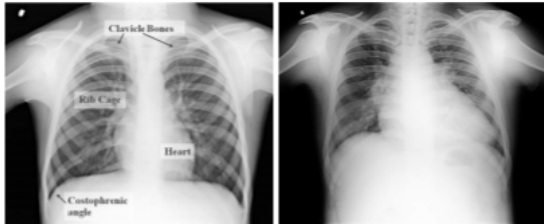

Fig 3.1 Anatomical features in chest x-ray and its variability

Image segmentation is very important for detecting good features. The functions of image segmentation are grouping pixels of same intensity from whole image, separate desired regions from original image and eroding unwanted regions.

Jaime Melendez, et al, [1] proposed a novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-Rays. In this work, Computer-Aided Detection (CAD) systems are typically optimized following a supervised learning approach that relies on large training databases comprising manually annotated lesions. CAD systems use the computer systems that aids in the creation, modification, analysis, or optimization of a design. Manually outlining those lesions constitutes a difficult and time-consuming process that renders detailed annotated data. Hence multiple instance learning (MIL) is used for more optimization.. This system is based on miSVM (multiple instance learning using support vector machines) technique. Jaime proposed an improved algorithm that overcomes miSVM's drawbacks such as underestimation of positive instances and iteration costs. It has two stages namely texture feature extraction and pixel classification.

- Texture Feature Extraction- Features are extracted based on the first four moments of the intensity distributions on the input image. The moments are computed by placing a grid in the lung field. The texture features obtained are normalized vertical and horizontal positions, distance to lung wall, centre of gravity of both lungs etc. are obtained.

- Pixel Classification- It is done using k nearest neighbor classifier. Feature vectors corresponding to normal pixels in abnormal images are ignored. After classification a heat map indicating abnormality is produced.

The limitations of this work include the following:
1) Manually annotating large x-ray data set is a tedious and a time-consuming task.
2) Diffuse aspect of lesions is not easy to be accurately outlined.
3) Disagreement between readers might occur, producing poor results.
4) If new sample data is available, training process is to be carried out again.
Segmentation of the lung field using CAD system is a difficult and major task because it is required for the proper analysis of chest X-Rays. Using this method, it is difficult for X-Ray analysis due to above limitations.

Candemir, et al, [2] presented a graph cut based lung segmentation method that detects the lungs. It is carried out in two stages –Average lung shape model calculation and lung boundary detection based on graph cut. First, it computes an average lung shape model using a set of training images. Then a graph cut based segmentation algorithm is used for detection of

lung boundaries based on the calculated model. An approximate shape model is computed by selecting set of training images based on shape similarity. Average of all selected mask is the lung shape model. Second, lung boundary detection is based on graph cut algorithm. The detection of lung boundary in X-Ray images is using image properties and lung shape model. Here the max flow min cut algorithm minimizes the objective functions to compute the global binary segmented image. i.e. global minimum which results in foreground and background labeling of pixels. This method is not accurate and has some robustness because a static shape model is not sufficient to describe the lung regions. Challenges faced are poor contrast images and images with different lightning conditions do not work well.

Stefan Jegar, et al, [3] presented a system for automatic detection of tuberculosis where lung field is segmented using a combination of masks- intensity masks, statistical lung model mask and a log Gabor mask. For a given input X-Ray, the lung field is segmented and a set of features are extracted for classification. The intensity mask is simply the complement of an X-Ray. It features on the dark parts of an image which are more likely to be part of the lungs. The lung model mask is a probabilistic lung shape model that is computed using the Japanese Society of Radiological Technology (JSRT) data set. The Log Gabor mask is based on Log Gabor Wavelets. It is a complex and true binary mask. Log Gabor filters are used to obtain larger spectral information while maintaining maximum spatial localization. The choice of selecting log Gabor filters is due to two main limitations. The maximum bandwidth of Gabor filter is limited and is not optimal if one is seeking for large spectral information with maximum spatial localization. Hence log gabor filter is used. Lung segmentation is performed in three steps. In the first step, a filter bank of 2 scales and 18 orientations is applied to the input X-Ray. The image is scanned from 0° to 180° in 10° steps. After log Gabor filtering, morphological operations are applied on the image. The lung fields can be seen as two large regions close to the center of a chest X-Ray. In the third step, regions near the center of the image are searched. The overall lung segmentation mask is the average of the intensity mask, the lung model mask and the log Gabor

mask. For classification, following features are taken into consideration- intensity histograms, gradient magnitude histograms, a shape descriptor, a curvature descriptor, angle between the x-axis and largest eigen vectors, histogram of oriented gradients, local binary patterns. Each descriptor is represented using 32 histogram bins and hence overall number of features is 7*32 = 224 features. Using the features, linear support vector machines is trained and classifies a given input image into either normal or abnormal. The main feature of this approach is that it combine different mask to achieve superior lung segmentation. This method creates a static lung model which provides less accuracy and even robustness.

The work done by Kim Le [4] is based on Watershed segmentation. Watershed segmentation is used to partition the area into regions. A gray scale image can be considered as a topographic surface. Higher intensity denotes peaks and hills and lower intensity denoted valleys. The water is initially filled in isolated valleys (local minima) using different color water (labels). Different color water tends to merge. To avoid this, barriers are built in the location where water merges. The filling of water and barrier continues until all the peaks are under water. The barriers created give the segmentation result. This is the concept of watershed segmentation. The segmentation process starts from a regional minima, M_i, the lowest points in the area into which the water can flow. With an appropriate distance measure, the area is divided into Ω_i regions. A point is added to a region, if the distance of that point from the region is smaller than those from other regions. The addition is repeated until no more point can be assigned to any region. The remaining unlabelled points are those in the watershed line. To obtain a thin line, a point is added to the region, Ω_i even if the distance of that point from the region equals those from other regions. A modification of watershed segmentation is a water snake approach which is used to adjust the smoothness of region boundaries. Here a new measure, gravity measure is introduced to segment a lung x-ray image.

Van Ginnekan, et al, [5] proposed an automatic detection of abnormalities in chest radiographs using local texture analysis. In this system, Active Shape Model (ASM) segmentation is used. Here corresponding

regions within the lung field is to be found. This is a hybrid method that combines rule based scheme with a pixel classifier. Here 42 regions are extracted from the lung field using ASM. A multi scale filter bank is used to extract the texture features. The filters are based on Gaussian derivatives. Texture Analysis creates a texture feature vector for each region in an image. To find abnormality in each region, a training dataset is produced. This set consists of feature vectors of abnormal regions and also equal number of feature vectors of normal regions. k- Nearest Neighbour (k-NN) classifier is used for classifying regions to normal or abnormal. The classification is a number in the range from zero (normal) to one (abnormal). In this system segmentation is most accurate at the vertical parts of the rib cage. Failures occur when detecting the diaphragm and lung tops. This method of segmentation is not suitable in areas that require high accuracy in terms of lung boundary.

Another related work of automated tuberculosis detection system was put forward by Binu Joykutty, et al[6], where segmentation process is done based on adaptive thresholding. In this, the main three steps in detection process are lung segmentation, feature extraction and classification. For the given input X- Ray image, lung region is segmented using adaptive thresholding. Next process is to extract the features followed by third stage of classification which outputs its decision whether the input chest image is normal or abnormal. The advantage of using adaptive thresholding is its capacity to change the threshold dynamically over the image. The inaccuracy caused by varying segmentation performance under different lighting conditions in Candemier's [2] graph cut based segmentation can be overcome using adaptive thresholding. Before segmentation, a mask is prepared which includes preprocessing such as contrast enhancement, erosion, inversion, border clearing, small objects removal and finally dilation. The segmented lung image is obtained by multiplying the mask with its original image. The second stage is to extract relevant set of features for the detection purpose. The feature set is referred as object detection inspired feature set .Here six features are extracted for detection purpose. For classification, k-NN classifier is used to classify normal/abnormal images. The advantage of this system is its simple implementation and added performance and accuracy compared to graph cut based segmentation

IV. PROPOSED SYSTEM

The proposed system can be explained in 3 stages.
- Lung Segmentation
- Feature Extraction
- Classification

For the given input chest X- Ray, lung region is segmented using adaptive thresholding followed by Expectation – Maximization (EM) segmentation method getting the desired region of interest. Then further features are extracted followed by the third stage of classification which outputs its decision whether the input chest image is normal or abnormal. A trained set is also provided for classification.

A. *Lung Segmentation*

The input is always a grayscale or color image and the output generated is a segmented binary image. The segmented lung image is obtained by multiplying mask with its original image and then it proceeds to feature extraction. Initially, a mask is created for segmentation. The mask preparation includes several pre processing steps. They are contrast enhancement, erosion, inversion, border clearing, small objects removal and finally dilation. Fig 4.1 shows the various pre- processing steps.
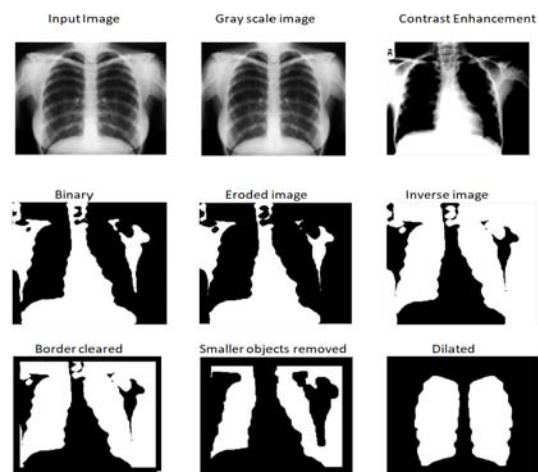

Fig. 4.1 Mask Preparation

Segmentation is done using adaptive thresholding. The steps involved are as follows:
1. Input the image to be segmented.
2. Input the default parameter values

a. Specify the size of the window over which the local weighted mean or median is computed.

b. Specify the filter type: Gaussian filter or median filter.

- If filter type is 'Gaussian', it performs Gaussian smoothing to obtain local weighted mean
- If filter type is 'Median', it performs median filtering to obtain local reference value for setting the threshold.

c. Set the threshold mode which specifies the way the threshold is defined. It can be fixed or relative.

- If threshold mode is 'relative', the value of 'c' represents the percentage relative to the local average gray value.
- If threshold mode is 'fixed', the value of 'c' represents the fixed level relative to the local average gray value, below which the local threshold is set.

d. Set a constant parameter 'c' as 15. This parameter is expressed either as a percentage or as a fixed value relative to the local average or median value depending upon the value of threshold mode used. Default value is 15.

3. Apply Gaussian or median smoothing denoted as fim to find local weighted mean

4. Apply threshold which is the difference between the weighted average and constant parameter 'c'

$$T= fim-c$$

5. Comparing threshold value obtained with the pixel value at each region of the image.

- If pixel value $\geq$ threshold value, it is set to foreground value.
- If pixel value $<$ threshold value, it is set to background value.

Fig. 4.2 shows an original chest X- Ray image and its region of interest after segmentation. The segmented lung image is obtained by multiplying mask with its original image.

Fig. 4.2 Lung Segmentation using adaptive thresholding.

For better segmentation results, EM segmentation is done. Expectation Maximization algorithm is implemented here. To understand EM segmentation, the concept of clustering is explained. Clustering is grouping of pixels into regions in images. Clustering can be divided into two namely Hard clustering and Soft clustering. In hard clustering, data point or elements either belong to a cluster or do not belong to a cluster but in soft clustering clusters may overlap. In this case, data points belong to multiple clusters at the same time but with different degree of belief. Clustering, specifically soft clustering can be represented using mixture models. Here each cluster corresponds to a generative model. Each cluster can be represented using a probabilistic distribution (Gaussian/ multinomial). To generate these models, we need to discover certain parameters like mean, covariance etc. EM algorithm automatically discovers all parameters for k- sources.

Consider data points $x_1 \ldots x_n$. Assume two Gaussian sources, yellow and blue. In this case, the source of the points is known. Thus mean and variance is given by,

$$\mu = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\Sigma = \frac{(x_1 - \mu_1)^2 + \ldots + (x_n - \mu_n)^2}{n}$$

Where n is the number of pixels in each cluster.

Suppose, if the source of the points is unknown, then the only way is to find the probability of presence of data points in either of the clusters. This is known as the chicken and egg problem. EM algorithm finds the probability of distribution of data points in each cluster.

1. Start with two randomly placed Gaussians, yellow (a) and blue (b)
2. (μ_a, σ_a^2) , (μ_b, σ_b^2) are the corresponding mean and covariances of the two clusters, yellow and blue respectively.
3. For each point, calculate the probability of presence in each cluster.
4. Once probability is found, readjust the mean and variance of the clusters to fit the points assigned to them.
5. Iterate steps 3 and 4 until convergence.

B. Feature Extraction
The second stage of feature extraction is performed on the segmented lung image. It

modifies the input image into some relevant set of features or information. For the segmented region, object detection inspired feature set is used here to extract vital features that helps in taking fine, delicate structures from chest radiographs and improves the detection process.
Object detection inspired feature set: It is a resourceful feature set that is enclosed by shape, edge and texture descriptors. Each feature descriptor is analyzed using a histogram which shows the various descriptor values across the lung field. The following are the features extracted for detection purpose [4]:

• Intensity histogram (IH)
• Gradient magnitude histogram (GM)
• Shape descriptor histogram (SD)
• Curvature descriptor histogram (CD)
• Histogram of oriented gradients (HOG)
• Local binary patterns (LBP)

1. Intensity Histogram: In this, histogram is based on various intensity descriptor values across the lung field. It is found that compared to pixels in the surrounding tissues pixels inside the lung fields will have lower values. In an image processing context, the histogram of an image normally refers to a histogram of the pixel intensity values. The histogram of an image is a graph showing the number of pixels in an image at each different intensity value found in that image. For a 8 bit grayscale image, there are 256 different possible intensities and so the histogram will graphically display 256 numbers. Here image is scanned in a single pass and a running count of the number of pixels at each intensity value is kept.

2. Gradient Magnitude Histogram: As the name suggests, it represents the magnitude of gradients in a histogram. Gradient can be defined as the directional change in the intensity of an image. Gradient is calculated by taking first order derivatives along X and Y directions. In edge pixels, gradient values will be higher.

3. Shape Descriptor Histogram: In this descriptor, the lung region is divided into sectors. Here the value in each bin corresponds to area of each sector.
Shape Descriptor is given by,

$$SD = \tan^{-1}(\lambda_1 / \lambda_2)$$

where $\lambda_{1 \text{ and }} \lambda_2$ are eigen values of Hessian matrix.

4. Curvature Descriptor Histogram: Histogram is built based on curvature properties. It is defined by,

$$CD = \tan^{-1}\left(\frac{\sqrt{\lambda_1^2 + \lambda_2^2}}{1 + I(X,Y)}\right)$$

CD ranges from $0 \leq CD \leq \pi/2$. I(x,y) denotes the pixel intensity. $\lambda_{1 \text{ and }} \lambda_2$ are eigen values of Hessian matrix, with $\lambda_1 \leq \lambda_2$

5. Histogram of oriented gradient: It refers to the distribution of direction of oriented gradients. Steps involved are pre-processing, gradient calculation, histogram of gradient calculation, normalization, and then calculate feature vector. The segmented lung image is divided into small cells. The gradient calculation is done for each cell followed by histogram of gradient is calculated. The descriptor is defined by the sequence of histograms computed for each cell.

6. Local binary pattern: This is a texture descriptor. It identifies the edges by measuring the average intensity between centre pixel and neighbouring pixels. Its simple implementation makes it an efficient texture descriptor.
All the feature descriptors in the figure are combined together to form a feature vector where each histogram bin corresponds to a feature. This feature is further given as input to the next stage of classification. Here, 32 bins are used for each histogram, and thus the total number of features is 6*32=192 which is enough to get feasible results.

C. Classification
k- NN classifier is used here to classify input image as abnormal/normal images. The analogy of k- Nearest Neighbours is 'Tell me who your friends are and I will tell you who you are'. The similarity between instances is identified by taking k- nearest neighbours. If majority represents abnormal pixels then it is classified as abnormal else normal. In trained set, abnormal images are labelled as 0 and normal images are labelled as 1. During classification, once a similar most sample is identified to the test sample, its label is assigned to the test sample. It outputs a value of 0 if the test image is abnormal; otherwise it returns a value of 1 to denote that the test image is normal.

k-NN classifier has several advantages over other classifiers. Some of them are

- Simple implementation
- No explicit training of data is required
- Computational complexity is less
- Less time taken for classification

For better performance during classification, k-NN is replaced by hybrid classifier. This is a classification technique that involves usually small number of heterogeneous methods that complement each other. The classification decision is reached by many different methods and finally combining their decisions. In this system, k- NN and Support Vector Machine (SVM) is used. It is a supervised machine learning algorithm that can be employed for both classification and regression purposes. It is based on the idea of finding a hyper plane that best divides a dataset into two classes. Support vectors are the data points nearest to the hyper plane. If the data points are far from the hyper plane, it is said to be correctly classified. When new testing data is added, whatever side of the hyper plane it lands will decide the class that we assign to it. Support Vector Machine is an efficient classifier because of its accuracy and also it uses a subset of training points for classification.

V RESULTS

The experiment is carried out on the chest X-ray images in the standard JSRT database which includes a total of 247 images, of which 154 are normal images and 93 are abnormal images. Both training and testing purposes are performed on this database.

The image is given as input and segmentation is done using adaptive thresholding and EM segmentation. Once the segmented lung image is obtained, feature extraction is performed to extract useful features. The derived feature set is given to hybrid classifier which classifies the image as abnormal or normal.

A. Analysis

The system is analysed for its performance based on Receiver Operating Characteristics curve (ROC) and accuracy. The quality of a classifier is determined by its classification performance. ROC curve is constructed to graphically represent the quality of a classifier. Based on threshold, different operating points can be obtained from a ROC curve. It is created by plotting the true positive rate (TPR) against the false positive rate.

Sensitivity and specificity are two performance measures for a binary classification test. Sensitivity (True positive rate) measures the proportion of positives that are correctly identified. In this case, people who are correctly identified as having the condition. Specificity (True Negative Rate) measures the proportion of negatives that are correctly identified as such. In this case, people who are correctly identified as normal. A False positive rate measures the proportion of positives that are wrongly identified. In this case, people who are wrongly identified with the condition when they does not have one. False positive rate is calculated as 1-specificity. A false negative rate measures the proportion of negatives that are wrongly identified. In this case, people who are wrongly identified as normal when they actually have the condition.

Accuracy indicates how accurate the test in reducing misclassification rate and is given by

$$\text{Accuracy} = \frac{a+b}{Total\ samples} * 100\%$$

where a and b are true positive and false positive respectively.

Sensitivity is given by ,

$$\text{Sensitivity} = \frac{a}{a+b} * 100\%$$

where a and b are true positive and false positive respectively.

Specificity is given by,

$$\text{Specificity} = \frac{d}{c+d} * 100\%$$

where c and d are false positive and true negative respectively.

Table I represents confusion matrix of k- NN classifier. Out of 35 abnormal images, it correctly identified 28 abnormal images and from a total of 35 normal images, it correctly picked up 25 images.

TABLE I
CONFUSION MATRIX OF K- NN CLASSIFIER

| Test	Measures	
Positive	28(True Positive)	10(False Positive)
Negative	7(False Negative)	25(True Negative)

Table II represents confusion matrix of hybrid classifier. Out of 35 abnormal images, it correctly identified 30 abnormal images and

from a total of 35 normal images, it correctly picked up 32 images.

TABLE II
CONFUSION MATRIX OF HYBRID CLASSIFIER

Test	Measures	
Positive	32(True Positive)	5(False Positive)
Negative	3(False Negative)	30(True Negative)

The proposed system achieves an accuracy rate of 88.57%. The system achieved a sensitivity of 91.42% and specificity of 85.71%. The Area Under Curve (AUC) is a measure of how well a parameter can distinguish between normal or abnormal images. It can be obtained from the ROC Curve. It is obtained by plotting true positive rate against false positive rate. Fig 5.1 shows the ROC curve of k- NN and hybrid classifier.
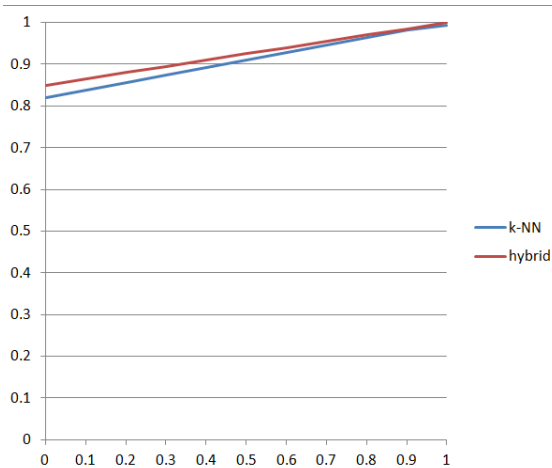
FPR- False Positive Rate   TPR- True Positive Rate
Fig 5.1 ROC Curve of k- NN and hybrid classifier

The Area Under Curve while using k- NN classifier is 90.86% and while using hybrid classifier is 92.78%.

VI CONCLUSION

This is a low cost computer aided detection system for tuberculosis from chest radiographs. There are mainly three detection stages namely image segmentation, feature extraction, classification. The segmentation is done using adaptive thresholding followed by EM segmentation. In feature extraction, six features are extracted. These features are called object detection inspired feature set. k- NN classifier is replaced by hybrid classifier to get better classification results. The accuracy came up to 88.57% and sensitivity 91.42%. The AUC for k-NN classifier is 90.78% and hybrid is 92.78%. This can be implemented in remote areas because of its simple implementation. The results show that hybrid classifier out performs k- NN classifier in terms of accuracy, sensitivity, specificity and area under ROC curve (AUC). Thus the results conclude that it is comparatively efficient than the existing primitive methods.

REFERENCES

[1] Jaime Melendez, et al, "A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest x-Rays ", IEEE Transactions on Medical Imaging, vol. 34, no. 1, January 2015, pp.179-192.

[2] S. Candemir, S. Jaeger, K. Palaniappan, S. Antani, and G. Thoma, "Graph-cut based automatic lung boundary detection in chest radiographs", IEEE Healthcare Technology Conference: Translational Engineering in Health and Medicine, 2012, pp. 31–34.

[3] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks", International Conference IEEE Engineering in Medicine and Biology Society (EMBS), 2012, pp. 4978– 4981.

[4] KIM LE,"Automated Detection of Early Lung Cancer and Tuberculosis Based on X-Ray Image Analysis", Proceedings of the 6th WSEAS International Conference on Signal, Speech and Image Processing, Lisbon, Portugal, September 22-24, 2006

[5] B. van Ginneken, S. Katsuragawa, B. ter Haar Romeny, K. Doi and M.Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis", IEEE Transactions on Medical Imaging, vol. 21, no. 2, Feb. 2002, pp. 139–149.

[6] Binu Joykutty, K. G Satheeshkumar, Binoshi Samuvel, "Automatic Tuberculosis Detection using Adaptive Thresholding in Chest Radiographs", International Conference on Emerging Technological Trends [ICETT],October 2016.