Computer Assisted X-Ray Analysis System for Detection of Onset of Tuberculosis

Shafeena Basheer, Jayakrishna V, Akas G Kamal

Abstract— This paper presents an initial analysis of the chest radiograph for the detection of pulmonary tuberculo-sis(PTB).In this, we propose to collaborate with the clinicians in National Institute of Research in Tuberculosis (NIRT) to come out with a software system to determine whether a patient has Tuberculosis or not, using image processing techniques applied to his digital x-ray image.

Index Terms— Pulmonary Tuberculosis, Digital chest X-ray,CAD,NIRT.

————————————— ◆ —————————————

I. INTRODUCTION

Automated analysis of Electronic Medical Records is a virgin research area where clinicians look forward to computer science to enhance the speed of arriving at correct diagnosis. This is of great potential, especially for rural areas in the third world countries, where the ratio of number of doctors to the total population is much below recommended standards. With automated analysis, a part of the analysis routinely performed by doctors could be delegated to computer software, thereby enabling a paramedic or field worker with a tablet PC connected with medical instrumentation to determine the onset of a disease. This will partially relieve the load from the doctors. The introduction of radiography as a diagnostic tool was a landmark in our knowledge of the natural history of tuberculosis and its diagnosis. It is still widely believed that pulmonary Tuberculosis can be diagnosed by chest radiography alone. Given the shortage of radiologists compared to a huge amount of chest radiographs or chest X-ray (CXR) images to be examined, a computer-aided detection/diagnosis (CAD) system is necessary to reduce the work of radiologists. On the other hand, tuberculosis (TB),especially infectious TB, such as post-pulmonary(reactivation) TB and HIV-related TB continues to be a public health problem of global proportions, especially in developing countries. Although chest radiography is increasingly important in the fight against TB, the sensitivity 70% 80% and specificity 60% 70% of radiologists diagnosis using conventional chest radiography are low and inter-reader

variation is high[1]. CAD systems can pave the way for the automated or semi-automated diagnostic interpretation of chest radiographs.Current studies shows that over one third of the world's population is infected with TB.Among this around 2.9 million people die every year from Pulmonary TB(Microbiology TexTBook, 2005).Around 60 people be-come infected for every single second.If this rate remains the same,over the next few years a large million of people could die from Tuberculosis.There have been many tech-nical and image processing advances since the discovery of X-rays over a century ago; screening for TB and other lung processes on chest X-rays has lagged behind until very recently. Perhaps the landmark event that will lead to the largest public health de-velopment in this important area is the transition of film based systems to digitized radiography.During the epidemiological surveys (Figure1)on large population more than 1000s of x-rays where taken from peoples in that area.During this mobile screen-ing program as defined by the WHO[2],all participants who either have CXRs suspicious for TB or whose X-rays are minimally abnormal should be intentionally over-read by the interpreter, and have an on-the-spot sputum collection[3].So this will results a false analysis of the number of tuberculosis patients.So,in our proposed method is to overcome the diffi-culties in screening the chest x-rays during the epidemiologi-cal surveys, by developing a system which is of good perfor-mance and less complex[3].

2. RELATED WORK

There have been many technical papers for the identifi-cation of nodules in lungs,detection of PTB from chest radiograph ,detection of cavities in lungs are proposed in the litera-ture.Among these some of the papers are of much importance

———————————————————

- *Shafeena Basheer:She is currently pursuing masters of Technology in computer science and engineering in MG University, India, E-mail:miz.shafee@gmail.com*
- *Jayakrishana V: He is assistant professor in computer science department in Amal Jyothi College of Engineering.E-mail:jayakrishnav@gmail.com*
- *Akas G Kamal:Professor in Adi Shankara Institute of Engineering and Technology,Kalady,India E-mail: akashgkamal@gmail.com*

for our proposed system.In [4],they uses a wavelet based tranform which is used to decom-pose the chest x-rays to high and low frequency compo-nents.Then the line profiles where taken and represented by daubechies co-efficients.And then a statistical analysis is used to test whether these features can identify TB.In [5],the authors propose a statistical interpretation method for detecting pulmonary TB from Chest X-rays.It is done by applying a wavelet transformation on the chest x-ray and calculating twelve texture measures from wavelet co-efficients.After this they apply a Principle Component Analysis(PCA)on the chest x-ray and then they find out misclassififcation probability by using probability ellipsoids and discriminant functions.

Figure 1. Chest X-rays Taken During Epidemiological Surveys

In [6],a phase congruency based method is used to classify all the pixels in the chest x-ray and then 4 statistical measures such as average, variance,coefficient of variation, and maximum PC value. In[7], authors uses a method based on Bayesian classi-fication to detects TB cavities in Chest X-Rays.Then to detect the shape of potential cavities,a circularity measure and to describes textures a gradient inverse coefficient of variation (GICOV) are used.There are also a number of other techniques where proposed by different authors.

III.MATERIALS AND METHODS

A. Database

This study involved in collaboration with National Institute of Research in Tuberculosis(NIRT), Chennai. NIRT is a premier institute under the Indian Council of Medical Research (ICMR),is an internationally recognized organisation for Tu-

berculosis (TB) research. Cases that arrived at the NIRT may be considered a random sample since an individual case may come from any of the Chennai hospitals or sub centers. Patients come to NIRT are advised to take the chest X-ray for immediate diagnosis. In NIRT, all pulmonologist are trained to interpret chest radiographs. There is a DICOM facility available at NIRT so that the doctors can view the X-rays on their PC.But during epidemiological surveys(figure 1) the doctors need to screen 1000s of x-rays .It takes weeks and even months.So to make this screening effortless we propose our method for the identification of onset of TB from Chest X-Rays.The database was created by collecting the chest x-rays from NIRT,chennai.The patient's chests X-ray are then divided into two sets which are the train set and the test set. The selected patients used as the train set (PTB present) were the confirmed PTB cases with no other systemic diseases such as diabetes, hypertension and heart disease. The confirmation of the PTB cases is based on the clinical feature (symptoms and sign), chest X-ray examination, and sputum Acid Fast Bacilli (AFB) direct smear. For PTB absent cases, normal lung (NL) of healthy individuals chest X-ray films selected by the NIRT represent contacts who came along with patients. The train set consists of 30 PTB present cases and 30 PTB absent cases. The test set consists of 20 PTB present cases and 20 PTB absent cases.During the process of training 70 % of the images from the traing set is used for feature extraction and classification and the remaining 30% is used to test the features and classification algorithm.

B. Digital X-Rays

The imaging features of active TB and inactive disease do have some unique features, but also overlap. Within the lung, imaging features of active pulmonary TB include but are not limited to the following manifestations:

- Cavity formation, a finding in the lung with a detectible radio dense rim.
- Air space consolidation small or large, that is segmental or lobar opacity in the lung.
- Miliary pattern is a fine granular sandy or seed-like appearance throughout the entirety of both lungs.
- Bronchiectasis or enlargement of airways can appear as tubular rings or cylinders of irregular diameter .

Figure 2. Digitized Chest X-ray image of a healthy Individual with normal Lung,(Source: National Institute of Research in Tuberculosis ,Chennai

The Figure 2 shows the Digitized Chest X-ray image of a healthy Individual with normal Lung and Figure 3 Digitized Chest X-ray image of a TB patient.
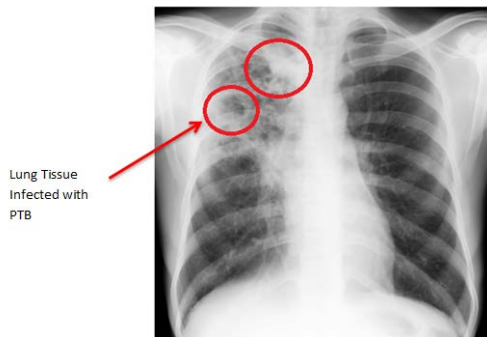

Lung Tissue Infected with PTB

Figure 3. Digitized Chest X-ray image of an Individual with Abnormal Lung,(Source: National Institute of Research in Tuberculosis ,Chennai

C. Pre-processing

The main goal of the pre-processing is to improve the image quality to make it ready to further process-ing by removing or reducing the unrelated parts in the background of the x-ray images.It will prepare the x-ray for the next two-process segmentation and feature extraction. The noise and high frequency components removed by filters[15].Pre-processing methods use a small neighborhood of a pixel in an input image to get a new brightness value in the output image. Such pre-processing operations are also called filtration. There are obvious reasons for the need of image pre-processing:

- Improvement of image quality to meet the requirements of physician
- Noise reduction
- Contrast enhancement
- Correction of missing or wrong pixel values
- Optimal preparation of data for postprocessing Elimination of acquisition-specific artifacts

D. Segmentation

The goal of image segmentation is to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces,objects, or natural parts of objects.A segmentation could be used for object recognition, occlu-sion boundary estimation within motion or stereo systems, image compression,image editing, or image database look-up[8].Segmentation algorithms generally are based on one of 2 basis properties of intensity values.

- Discontinuity : to partition an image based on sharp changes in intensity.
- Similarity : to partition an image into regions that are similar according to a set of predefined criteria.

E. Feature Extraction

Features are the representatives of the images.In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction.When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features (also named features vector)[4]. Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input[16].The issue of choosing the features to be extracted should be guided by the following concerns:

- All features should carry enough information about the image.
- All the features should be easy to compute.
- They should relate well with the human perceptual characteristics.

F. Classification

Image classification refers to the task of extracting information classes from a large images. Depending on the interaction between the analyst and the computer during classification, there are two types of classification: supervised and unsupervised.

- Supervised classification: Supervised classification uses the spectral signatures obtained from training samples to classify an image[9].
- Unsupervised classification:Unsupervised classification finds spectral classes (or clusters) in a multiband image without the analysts intervention.

IV. GENERAL PLAN OF IMPLEMENTATION

The general plan of implementation consists of three phases.These are:

- Generation of a large corpus of annotated test cases. This will contain sufficient number of diverse records with the disease in various stages and also without

the disease.

- Generation of an automated classification algorithm that can be shown to work well on the above corpus. This is the most crucial part of the development.
- Implementation of above algorithm in the operational environment in NITR and determining effectiveness.

A. Corpus Generation

Figure 4 gives the block diagram of the implementation plan for corpus generation. The input to this effort is the set of digital x-rays and stores into the corpus. The corpus as above has to be padded with meta data, for access by subsequent analysis routines. For this, a metadata generation program will be developed. This will allow a data entry operator to create a soft copy of the tiff image data along with required meta data taken from the clinicians notes in the case sheets. This will act as the test bed for all future analysis.
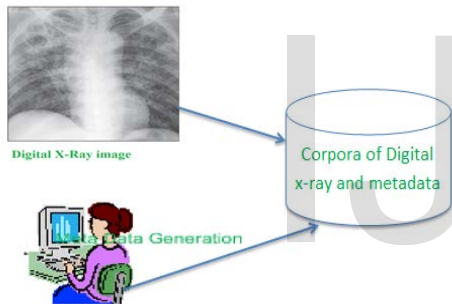

Figure 4. Overall Scheme of Corpora Generation

B. Generation of Automated Classification Algorithm

Figure 5 gives the block diagram of the plan for gen-eration of classification algorithm. We start with literature survey for identification of the pre-processing steps (image enhancement, segment extraction, etc.) prior to parameter-ization. These will be implemented on a representative set of images in the corpus and the outputs will be shown to the clinicians to obtain his feedback. After a few iterations, this will converge to a lo-cal optimum at which time we will proceed to the next step of determining the optimum set of parameters to be extracted from the pre-processed image.Arriving at the optimum set of parameters is an iterative process. We start with an initial set of parameters based on literature survey with extensive dis-cussion with the clinicians. These are extracted from a refer-ence set of PMRs that has equal number of cases with and

without tuberculosis. The reference set is thus partitioned into two disjoint sections. A principal component analysis will bring out the set of parameters contributing to the max-imum extent for this classification. The optimum linear discriminator with these parameters can be used as the classifier. The classi-fier is applied to the entire corpus and the error is computed. In case of large errors, we will have to re-visit the parameteri-zation scheme to de-termine whether alternate parameteriza-tion process could give better results. The discussions with clinicians will help in this. The process is to be continued till we get a stable classifier with acceptable error limits. This will enable a good classifier for the analysis of the x-ray images. Applying this classifier over the corpus will results classified x-rays over the corpus. Continuing this process untill we get an appropriate classification of x-rays over the corpus.
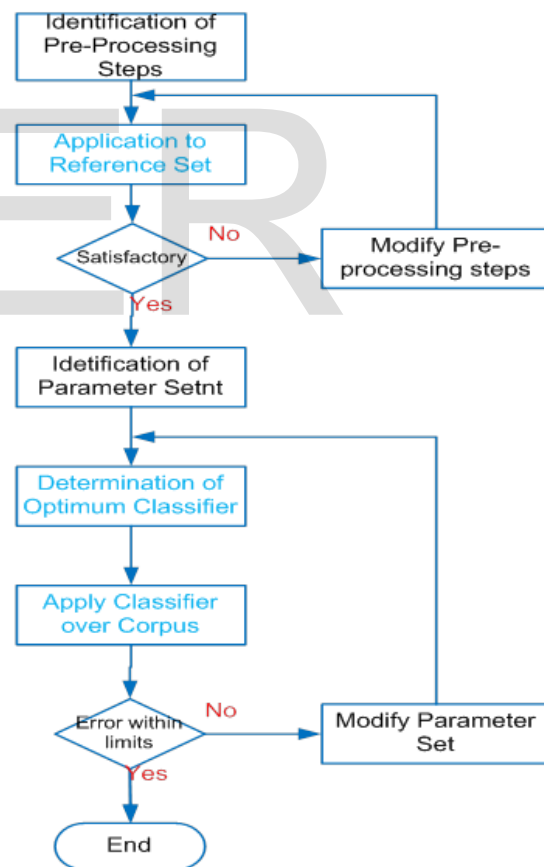

Figure 5. Generation of Classifier Algorithm

C. Implementation in Operational Environment

Figure 6 gives the flow graph of the working of the system in operational environment. The image from the digital x-ray

machine will be sent to the tablet PC or computing station which will perform the steps of pre-processing, parameterization and classification. The clas-sification result is included in the EMR for the benefit of consulting clinicians. In case they disagree with the computer generated classifier, it is indicated in the EMR comment section. These discrepancies are brought out periodically so that modifications to the classification schemes could be attempted later. Along with the discrep-ancy report, the classification efficiency also is brought out. This will help the doctors determine whether the proposed scheme is robust enough for usage in epidemiological surveys.
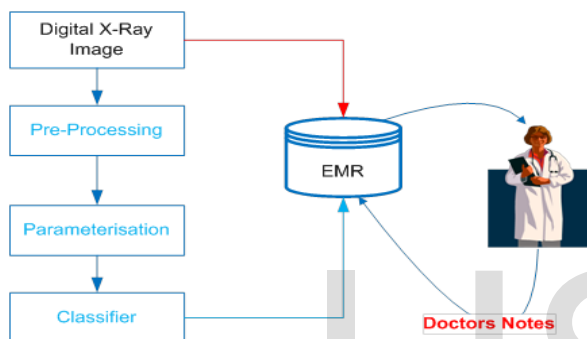

Figure 6. Operational Phase

V. PROPOSED SYSTEM

A. Pre-processing & Segmentation

1) **Pre-processing**: A median filter operates over a window by selecting the median intensity in the win-dow.Median filtering is a nonlinear operation often used in image processing to reduce noise. Such noise reduction is a typical pre-processing step to improve the results of later processing.The figure 7 shows the image of a PTB patient and the median fil-tered output.
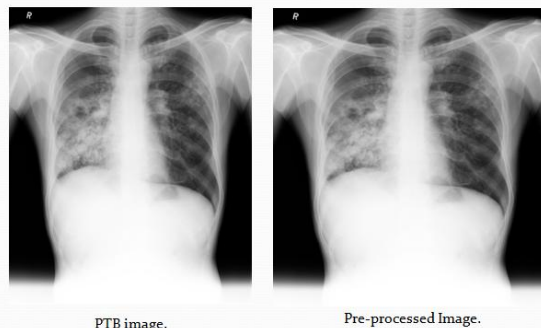

PTB image.   Pre-processed Image.

Figure 7. X-Ray image with Left Lower Zone affetected with PTB and Median Filterd output.

Segmentation: Segmentation algorithms are based on one of two basic properties of intensity values dis-continuity and similarity.Threshold is one of the widely methods used for image segmentation. It is useful in discriminating foreground from the background.Threshold techniques can be categorized into two classes: global threshold and local (adaptive) thresh-old. In the global threshold, a single threshold value is used in the whole image. In the local threshold, a threshold value is assigned to each pixel to determine whether it belongs to the fore-ground or the background pixel using local information around the pixel. Because of the advantage of simple and easy implementation, the global threshold has been a popular tech-nique in many years.By selecting an adequate threshold value T, the gray level image can be converted to binary image. The binary image should contain all of the essential information about the position and shape of the objects of interest.The ad-vantage of obtaining first a binary image is that it reduces the complexity of the data and simplifies the process of recogni-tion and classification. The most common way to convert a gray-level image to a binary image is to select a single thresh-old value (T). Then all the gray level values below this T will be classified as black (0), and those above T will be white (1). The segmentation problem becomes one of selecting the prop-er value for the threshold .Figure 8 shows the otsu thresholded output of the preprocessed x-ray image.

Otsu Thresholding Method: Based on a very simple idea: Find the threshold that minimizes the weighted within-class variance. This turns out to be the same as maximiz-ing the between-class variance. Oper-ates directly on the gray level histogram [e.g. 256 numbers, P(i)], so its fast.The otsu assuptions are histogram (and the image) are bimodal,No use of spatial ... ... any other notion of object structure,Assum ... tat... statistics, but can be modified ... to ... adap-tive ... xercises),Assumes u... f... uminat... licitly), so th... bimodal brightnes ... be... ior arises ... ect ap-pea... ce differences onl...

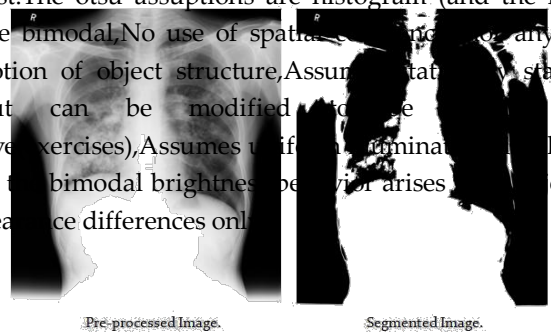Pre-processed Image.   Segmented Image.

Figure 8. Pre-processed x-ray image and it's thresholded output using

otsu method at level=.7

B. Feature Extraction & Classification

1) Feature Extraction: Scale-invariant feature trans-form (or SIFT)is an algorithm in computer vision to detect and describe local features in images[18].For any object in an image, interesting points on the object can be extracted to provide a feature description of the object.This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image are detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.So we use SIFT as the feature extraction technique for extracting the features from the segmented x-ray image[17].

2) Classification: We use PCA as the classification method for classifiying the x-ray images based on the features extracted by SIFT.PCA is a feature based clas-sification technique that is characteristically used for im-age recognition[16].PCA is based on principal features of an image and these features discreetly represent an image.PCA has ability to identify relatively fewer features or components that as a whole represent the full object state and hence are appropriately termed Principal Components. Thus, principal components extracted by PCA implicitly represent all the features[19]. However, these abstracted features may or may not include a specific feature[14].Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables[20].Principal features in PCA are represented by Eigenvectors. The Eigenvectors are defined to be a related set of spatial characteristics of an image that a comput-er uses to identify and recognize a specific cloud type. Eigenvectors of the covariance matrix is computed from the training set of images. These eigenvectors represent the principal components of the training images. These eigenvectors are often ortho-normal to each other[16].

VI. DISCUSSION

When we consider the creation of a CAD system we can think of a number pre-processing methods, parame-terization steps and classification methods. We are in the planning phase of this CAD system. We try to find out the best pre-processing method which is most suitable for making an X-ray high quality .And also we try to select a set of parameters which classify the images into different classes. After all we try to select a suitable classification algorithm to classify the normal x-ray images and PTB images.So with all these selection and effort we are now at a conclusion that by simply comparing the area we can initially identify whether there is any infections affected to lung,then we analyse some statistical properties[11] to compare the normal and PTB images.Some results are shown below.When comparing figure 9 and 10 we can see that the black portions in the affected lung shows less value than its non affected part.In normal images there is not much difference in the black area based on the count of black pixels.In figures 11 and 12 the results of statistical properties shows that for PTB images the values are different for different lung zones.This results is obtained by dividing the PTB image in the left part of figure 7 .Along with this simple analysis of chest x-ray ,we also do some feature extraction method based on SIFT and perform the classification.The classification on the dataset results that the images in the test set shows above 50 % correct classification.

Image	Left-Black Area	Right-Black Area
Image1	375060	393606
Image2	573475	574101
Image3	661219	668762
Image4	382730	351033
Image5	263531	229121

Figure 9. Area of left and right lung of 5 PTB images at thresholding level=.7

Image	Left –Black Area	Right-Black area
Image1	522219	290635
Image2	434020	2476
Image3	363871	628547
Image4	243729	369197
Image5	362842	175564

Figure 10. Area of left and right lungs of 5 Normal images at thresholding level=.7

PTB Image	Contrast	Standard Deviation	Homogeneity	Mean	Energy
Left Upper	0.0679	94.68392358	0.9661	72.26942048	0.1957
Left Lower	0.0583	113.88169054	0.9609	117.49990811	0.2881
Right Upper	0.0577	86.30977964	0.9712	33.91234092	0.2142
Right lower	0.0608	84.96106738	0.9796	56.79305400	0.2041

Figure 11. Statistical Property estimation of a PTB image at threshold-ing level=.7

Normal Image	Contrast	SD	Homogeneity	Mean	Energy
Left upper	0.4057	94.16634271	0.9905	71.36421251	0.6175
Left Lower	0.1653	90.60508051	0.9962	98.88863796	0.5137
Right Upper	0.2605	98.64512842	0.9937	61.20559575	0.6190
Right Lower	0.2265	94.60434248	0.9949	94.60315025	0.5067

Figure 12. Statistical Property estimation of a Normal image at threshold-ing level=.7

VII. CONCLUSION

This paper deals with the initial analysis of a CAD system for automated analysis of chest x-ray for identi-fication of pulmo-nary TB. The main focus of this paper is to give an introduc-tion for the concept of an automated identification of onset of TB. Lots of researches are going on in this area from several years. Based on all these researches and experiments we try to implement a system which can automatically detect the possi-bility of TB in a chest radiograph. This will considerably re-duce the effort of Medical officer and radiologist.

VIII. FUTURE WORKS

In future we need to find out the exact feature extrac-tion method and classification algorithm.So that we can classify the PTB images from Normal x-ray images.

ACKNOWLEDGEMENT

REFERENCES

[1] Virginia Department of Health D Tuberculosis Control Pro-gram TB Infection Fact Sheet.

[2] WHO. Global Tuberculosis Report. World Health Organiza-tion, 2012.

[3] Stefan Jaeger1, Alexandros Karargyris1, Sema Candemir1, Jenifer Siegelman2, Les Fo-lio3,SameerAntani1,GeorgeThoma1,Automatic screening for tuberculosis in chest radiographs: a Survey.

[4] Noor NM, Rijal OM, Fah CY. Wavelet as features for tuberculosis (MTB) using standard x-ray film images.

[5] Noor,Rijal, Ashari Yunus,Aziah A. Mahayiddin, Gan Chew Peng, S. A. R. Abu-Bakar ,A Statistical Interpretation of the ChestRadiograph for the Detection of Pulmonary Tubercu-losis.

[6] Omar Mohd Rijal, Member, IEEE, Hossien Ebrahimian, Norliza Mohd Noor, Senior Member, IEEE ,Determining Features for Discriminating PTB and Normal Lungs Using Phase Congruency Model.

[7] Rui Shen, Student Member, IEEE, Irene Cheng, Senior Member, IEEE, and Anup Basu, Senior Member, IEEE,A Hybrid Knowledge-Guided Detection Technique for Screen-ing of Infectious Pulmonary Tubercu-losis From Chest Ra-diographs.

[8] Suzuki K, Abe H, MacMahon H, et al. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network(MTANN).

[9] Shen R, Cheng I, Basu A. A Hybrid Knowledge Guided Detection Technique for Screening of Infectious Pulmonary Tuberculosis from Chest Radiographs.

[10] Xu T, Cheng I, Long R, et al. Novel coarse-to-fine dual scale tech-nique for tuberculosis cavity detection in chest radiographs.

[11] Shi Y, Shen D. Hierarchical shape statistical model for segmentation of lung fields in chest radiographs

[12]   wasuta renkjumnong,SVD and PCA in Image Processing.

[13] R. Ramani,Dr. N.Suthanthira Vanitha,The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images.

[14] Imran S. Bajwa1, M. Shahid Naweed, M. Nadim Asif, S. Irfan Hy-
      der,Feature Based Image Classification by using Principal Component
      Analysis.

[15] Daniela Stan Raicu,Image Feature Extraction.

[16] Tony Lindeberg, KTH Royal Institute of Technology, Stockholm,
      Sweden ,Scale Invariant Feature Transform.

[17] http://en.wikipedia.org/wiki.

[18] SUSANTA KUMAR SAHU,Image Segmentation for Text Extraction.
[19] Feature Based Image Classification by using Principal Component
      Analysismore by Imran Sarwar Bajwa.

IJSER