International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 4 Issue 7, July 2015

An Efficient Diagnosis Of Tuberculosis With The Aid Of Chest Radiographs

Anju Mathews, Jithin Jose Kallada

***Abstract*— Tuberculosis (TB) is considered as a major health threat in many regions of the world. It is considered to be the world's second leading cause of death. So it is necessary to diagnose TB before it becomes fatal. But diagnosis of Tuberculosis is still a challenge and when left undiagnosed and thus untreated, it is seen that the mortality rates of patients with tuberculosis are considerably high. There are various diagnostic methods which are put into practice and those diagnostic methods are developed in the last century is still considered as the standard diagnostics. But these diagnostic methods are unreliable and slow. In order to reduce the burden of the disease, this paper proposes an efficient diagnosis of tuberculosis with the aid of chest radiographs. Lung region is the most affected part of Tuberculosis. So we perform an K means algorithm to segment the lung region from the CXR image which is considered to be the existing work, which is less efficient, we don't get a clear cut of lungs. So we proposed lung segmentation using region growing segmentation. After performing lung segmentation we perform feature extraction, where various features of the lung region are extracted out. Of these features we consider the best feature mainly the HOG feature. The feature is then fed to a classifier mainly SVM classifier and this SVM classifier is then trained using images from the database and it then perform classification of the CXR image as whether TB is affected or not. Various level of TB affected is also shown according to the input CXR image. The TB affected area is also determined using region based active contour segmentation. The comparisons between various segmentations are also performed.**

***Index Terms*—Tuberculosis, K-means algorithm, Region growing algorithm, Active contour segmentation**

I. INTRODUCTION

Tuberculosis is the short for Tubercle Basillus and is considered as a major threat in many regions of the world. Earlier, it was called as Phthisis, consumption or Phthisis pulmonalis. It is a widespread infectious disease and in many cases fatal, which is caused by various strains of mycobacteria, mainly Mycobacterium Tuberculosis. TB mainly affects the lungs of individuals but it can also affect other parts of the human body. . It mainly spreads through air, mainly when people suffering from TB sneeze, cough or expel infectious bacteria when respiratory fluids are transmitted through the air. Infections which don't show symptoms are known as latent tuberculosis and about 25% of the people with TB are having latent TB.

TB is mostly seen in individuals in sub-Saharan Africa and Southeast Asia, and the main reason is poverty and malnutrition which reduces the resistance to the disease. There are various symptoms for detecting TB and it includes fever, chronic cough with blood tinged sputum, chest pain, night sweats, weight loss, fatigue, nail clubbing. Treatment of Tuberculosis is difficult but several antibiotics exist for treating TB. The technique for diagnosing TB is the identification of Mycobacterium tuberculosis in a pus sample, which is considered as the current standard. But the disadvantage with this method is that, it may take several months to identify this slow-growing organism in the lab. Another technique is sputum smear microscopy. In addition, there are various skin tests are available for detecting TB but the skin tests are not always reliable. Molecular diagnostic tests are the latest development for detection of TB which are fast, accurate, and highly sensitive and are specific too. However, for these tests, additional financial support is required. Thus for computed aided diagnosis based detection of various lung diseases, accurate, sharp segmentation of lung boundary is necessary.

In this paper, our work shows an automated approach for detecting TB in chest x-rays (CXRs) which are more accurate than the existing methods. A chest radiograph, which is colloquially called as a chest X-ray (CXR), or otherwise chest film. It is a projection radiograph of the chest which is used to diagnose conditions which affects the chest, its nearby structures and contents so that various diseases can be estimated. An automated technique of x-ray reading needs mass screening of large populations that could not be managed manually. For evaluating TB, a posteroanterior radiograph (x-ray) of a patient's chest is mandatory. Therefore we aim at a powerful TB diagnostics system, so a reliable screening system for TB detection using radiographs is needed. A CXR which is affected with TB will have cavitations, infiltrations, effusions, or military patterns in the x-ray images. Examples of normal CXR and abnormal CXR i.e., without and with TB infection are shown in fig.1 and fig.2 respectively. Fig.1 represents the normal CXR without any presence of TB, while fig.2 shows the abnormal CXR which shows the presence of TB in the X-ray in the form of cavities or nodules in the lung region.

Manuscript received July, 2015.
***Anju Mathews**, Department of Electronics and Communication, NCERC Pampady., Thrissur, Kerala, India.*
***Jithin Jose Kallada**, Department of Electronics and Communication, NCERC Pampady,Thrissur,Kerala,India.*

ISSN: 2278 – 1323

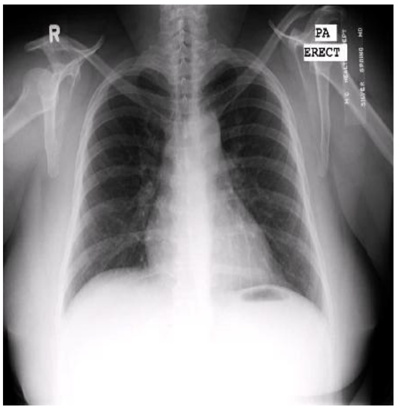

Fig. 1. Normal CXR

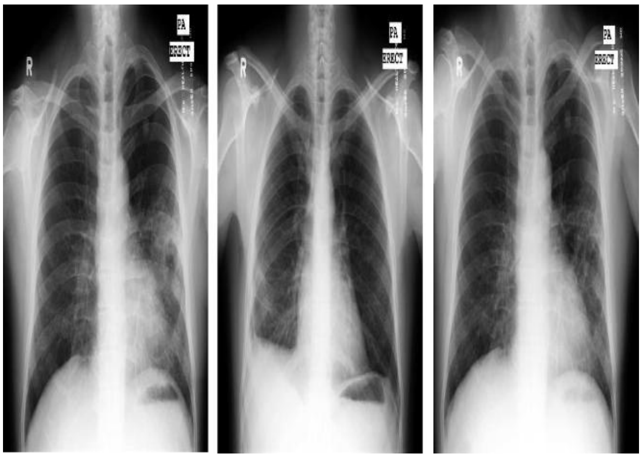

Fig. 2. Abnormal CXR's.
It is having nodules, infiltrates, cavities in the lung region.

II. SYSTEM OVERVIEW

This section presents the system overview and is shown in fig. 3. An automated approach for the detection of tuberculosis from CXR's is presented here and the steps include lung segmentation, followed by feature computation and classification of the input x-ray as TB positive or negative. At first, our system performs segmentation of the lung region of the input CXR using K means algorithm. Since the K means algorithm doesnot provide much accuracy in lung segmentation, we provide lung segmentation using Region growing algorithm. After lung segmentation, features are computed. Of which the best feature, HOG feature is computed and HOG feature is used for classification. We are using SVM classifier for classifying the Xray as TB infected or not.

So here we are employing segmentation using K means algorithm and also using region growing algorithm. Then using HOG classifier and SVM classifier, we classify whether it is TB affected or starting level TB or high level TB. WE also employ active contour algorithm for determining the area of TB infection.

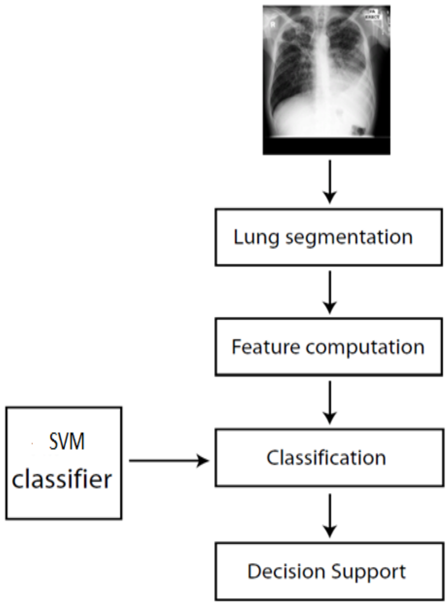

Fig. 3. System Overview

III. METHODOLOGY AND PROCEDURE

A. K means Algorithm

K means algorithm is considered to be a clustering algorithm whih is used to classify the input data points into various classes according to the inherent from each other. The algorithm is used to find natural clustering as it assumes that the data features form a vector space. The datapoints are clustered around centroids which are then obtained by minimizing the objective function.

K-Means is a least-squares partitioning method which divides a large collection of objects into K groups. The algorithm is as follows.

- For each cluster, we compute the mean.
- We compute the distance of each point from each cluster. It is done by computing its distance from the corresponding cluster mean. Then each datapoints are assigned to the nearest cluster.
- Repeat the above steps until the sum of squared within group errors cannot be lowered any more.

The datapoints are assigned to clusters initially is done randomly. Performing the iterations, the algorithm tries to minimize the sum, over all groups, of the squared within group errors, which are the distances of the points to the respective group means. Convergence is reached at a point when the objective function cannot be lowered any more. The groups which are obtained are geometrically as compact as possible around their respective means. Then a feature vector is constructed corresponding to each pixel using the set of feature images.

Once the image is segmented using K means algorithm, we can further improve the clustering by assuming that the neighboring pixels have a high probability that they fall into the same cluster. Thus, even if a pixel has been wrongly clustered, it can be corrected by looking at the neighboring pixels

B. Region Based Segmentation

The main aim of performing segmentation is to partition an image into regions. Region based segmentation is considered to be a technique which is used to determine the region directly. There are various other segmentation methods such as 'thresholding' and it performs segmentation by looking the boundaries between the regions and it is based on discontinuities in gray level or color properties.

Basic concept of seed points:

The first and foremost step in region growing algorithm is selection of a set of seed points. Seed point selection is according to user criterion. The initial region begins as the exact location of these seeds. Then the regions are then grown from the seed points to neighbouring points according to region criteria. The criterion can be gray level texture, colour etc. Various keypoints about region growing is as follows.

- The suitable selection of seed points is important.
- More information of the image is better.
- The value, "minimum area threshold".
- The value, "Similarity threshold value".

Advantages and disadvantages of region growing:

Advantages***:***

- With region growing, we can correctly separate the regions which have the same properties we define.
- It can provide good segmentation results which have clear edges.
- The concept of region growing is simple. First of all we only need a small numbers of seed point which represent the property we want and then grow the region.
- The seed point and the criteria can be determined according to the user.
- Multiple criteria can be chose at the same time.
- It can work well with noise.

Disadvantages:

- The computation is consuming, no matter the time or power.
- Noise or variation of intensity can cause result in holes or oversegmentation.
- This method may not distinguish the shading of the real images.

We can conquer the noise problem easily by using some mask to filter the holes or outlier. Therefore, the problem of noise actually does not exist.

Region Growing Methods:

There are a few points that should be taken into consideration when trying to segment an image. We must have regions that are disjoint because a single point cannot be contained in two different regions. The regions must span the entire image since each point has to belong to either one region or another. To get regions at all, we must have to define some property which will be true for each region that we define. In order to ensure that the regions are well defined and that they are indeed regions themselves and not several regions together or just a fraction of a single region, that property cannot be true for any combination of two or more regions. If these criteria are met, then the image is truly segmented into regions.

C. Region Based Active Contour Segmentation

Active contour model, also called snakes, is a framework in computer vision for delineating an object outline from a possibly noisy 2D image. This model is popular in computer vision, and is greatly used in various applications like object tracking, shape recognition, segmentation, edge detection and stereo matching.

A snake is an energy minimizing, deformable spline influenced by constraint and image forces and pull it towards object contours and internal forces that resist deformation. Snakes may be understood as a special case of the general technique of matching a deformable model to an image by means of energy minimization. Compared to classical feature attraction techniques, snakes have multiple advantages:

- They autonomously and adaptively search for the minimum state.
- External image forces act upon the snake in an intuitive manner.
- Incorporating Gaussian smoothing in the image energy function introduces scale sensitivity.
- They can be used to track dynamic objects.

The key drawbacks of the traditional snakes are

- They are sensitive to local minima states, which can be counteracted by simulated annealing techniques.
- Minute features are often ignored during energy minimization over the entire contour.
- Their accuracy depends on the convergence policy.

Region-based active models are known for robustness to weak edges and high computational

Active contours without edges have two force terms. The first term is the force to shrink the contour. The second term is the force to expand the contour. These two forces get balanced when the contour reaches the boundary of our interested object.

Region-based Active Contour Formulation:

The basic idea starts with an initial curve C, and the curve is then deformed to the boundary of the object, under some constraints from the image. The initial curve C for the contour which is then moved by image forces to the boundaries of the desired objects. Active contour algorithm attempts to minimize an energy associated to the current contour as a sum of an internal and external energy. The external energy is supposed to be minimal when the snake is at the object boundary position. The most straightforward approach consists in giving low values when the regularized gradient around the contour position reaches its peak value. The internal energy is supposed to be minimal when the snake has a shape which is supposed to be relevant considering the shape of the sought object. The most straightforward approach grants high energy to elongated contours (elastic force) and to bended/high curvature contours (rigid force), considering the shape should be as regular and smooth as possible.

Let $I: \Omega \rightarrow R$ is the image with labels extracted; we wish to find the boundary that separates Ω into the object region and background.

We make assumption that close to the boundary, the image is piecewise constant. That is, we consider an open curve C which is used to separate the local neighbourhoods into two regions, each with Gaussian statistics. We decompose Ω into

ISSN: 2278 – 1323

disjoint rectangles Bi, $\Omega = \cup^{k}_{i=1} Bi$ where we expect the intensity to be bi-modal Gaussian. Let $C_d{}^{o}$ denote the object region that lies within a Euclidean distance of d from C, similarly, $C^d{}_b$ denotes the background region that lies within a distance d from C. Also, $\mu_o = [\mu^o{}_1, \mu^o{}_2 \ldots \mu^o{}_K]$, $\sigma_o = [\sigma^o{}_1, \sigma^o{}_2 \ldots \sigma^o{}_K]$ gives the distribution in the regions $C^d{}_0 \cap Bi$ similarly μb, σb gives the distribution in $C^d{}_b \cap Bi$.

We minimize the following energy over the space of open, smooth parametric curves C: [0, 1] → Ω, distributions $\mu_o, \sigma_o, \mu_b, \sigma_b$.

$$E[C, (\mu^o, \sigma^o), (\mu^b, \sigma^b)] = \sum_{i=1}^{k} \int_{C^o_d} \chi B_i \left[\frac{(I-\mu^o{}_i)^2}{\sigma^{o2}} + \ln(\sigma^o{}_i)2 \right] dx + \sum_{i=1}^{k} \int_{C^o_d} \chi B_i \left[\frac{(I-\mu^b{}_i)^2}{\sigma^{o2}} + \ln(\sigma^b{}_i)2 \right] dx + \lambda \int_{c} ds \quad (1)$$

Given the estimates μo, σo, μb, σb, the first and second terms of the energy drives C to partition the rectangle Bi into two regions; $C^d{}_0 \cap Bi$ and $C^d{}_b \cap Bi$ where the distribution is close to $(\mu^o i, \sigma^o i)$ and $(\mu^b i, \sigma^b{}_i)$. The last term minimizes the length of solutions C, governed by parameter λ.

Given an initial guess C0, we use the Euler Lagrange equations of (1) to iteratively solve for $(\mu^o, \sigma^o, \mu^b, \sigma^b)$, and C, using an explicit finite difference scheme. Since only rectangles Bi that intersect with C affect the update equations, we compute distributions only at such rectangles. Further, for optimal number of rectangles used, and for better distribution estimates, rectangles are centered at $C^n = [x^n{}_k, y^n{}_k]$ for uniformly placed, discrete curve points k, and iteration n.

In Region based segmentation, the initial step is to specify the initialization mask based on the region of interest. Final Step is the active contour force analysis. In the active contour force analysis the abnormalities are identified by means of external and internal forces. As the iteration increases the contour line moves closely to the affected region. Finally it identifies the abnormality region as a line evolving curve .By analyzing this curve we can easily identify the affected region .and accordingly we can make necessary decisions.

D. *HOG Features*

Histogram of Oriented Gradients (HOG) are feature descriptors which are used in computer vision and image processing for the aim of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of various other methods like edge orientation histograms, scale-invariant feature transform descriptors(SIFT), and shape contexts, but differs in the fact that it is computed on a dense grid of uniformly spaced cells and it uses overlapping local contrast normalization for improved accuracy. The main feature of Histogram of Oriented Gradient descriptors is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. These descriptors can be implemented by dividing the image into small connected regions, which is called as cells. Then for each cell we perform compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The combination of these histograms then represents the descriptor. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination or shadowing.

Histogram of oriented gradients (HOG) is a descriptor for gradient orientations weighted according to gradient magnitude. The image is divided into small connected regions, and for each region a histogram of gradient directions or edge orientations for the pixels within the region is computed. The combination of these histograms represents the descriptor.

E. *SVM Classifier*

Classifiers are used to detect abnormal CXRs with TB. For this we use a support vector machine (SVM), which classifies the computed feature vectors into either normal or abnormal i.e., whether the CXR is affected with Tb or not. The support vector machine (SVM) classifier is a binary classifier which looks for an optimal hyperplane as a decision function. Once trained on images containing some particular object, the SVM classifier can make decisions regarding the presence of an object in additional test images. Ideally, the feature vectors of abnormal CXRs will have a positive distance to the separating hyperplane, while the feature vectors of normal CXRs will have a negative distance. If the distance is larger, the more confident we are in the class label. Here SVM classifier is used to detect the presence of TB in chest X rays.

IV. EXPERIMENTAL RESULTS

The lung segmentation using K means algorithm is shown in fig.4

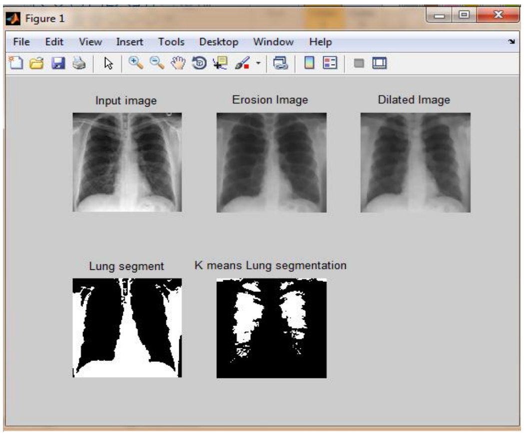

Fig. 4. Lung Segmentation using K means algorithm

The segmentation provided using Region growing algorithm is in fig.5

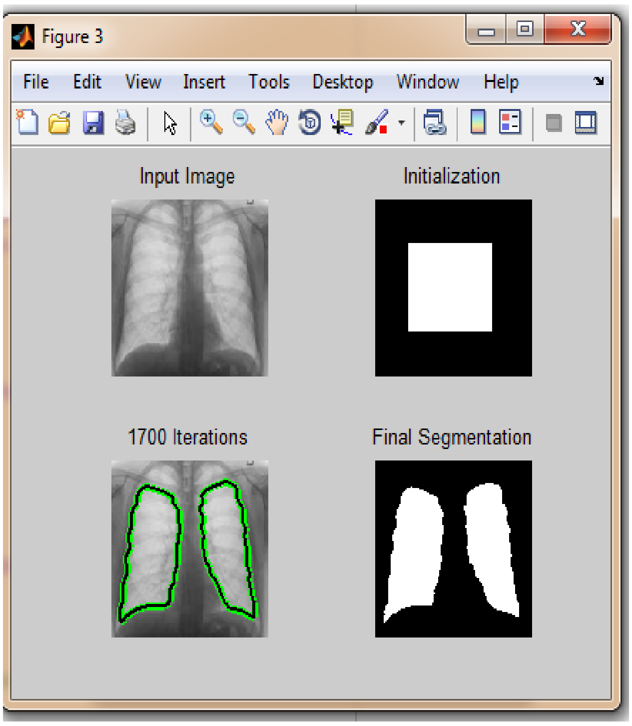

Fig. 5. Lung Segmentation using Region growing algorithm

The TB affected area is determined using active contour lung segmentation method.

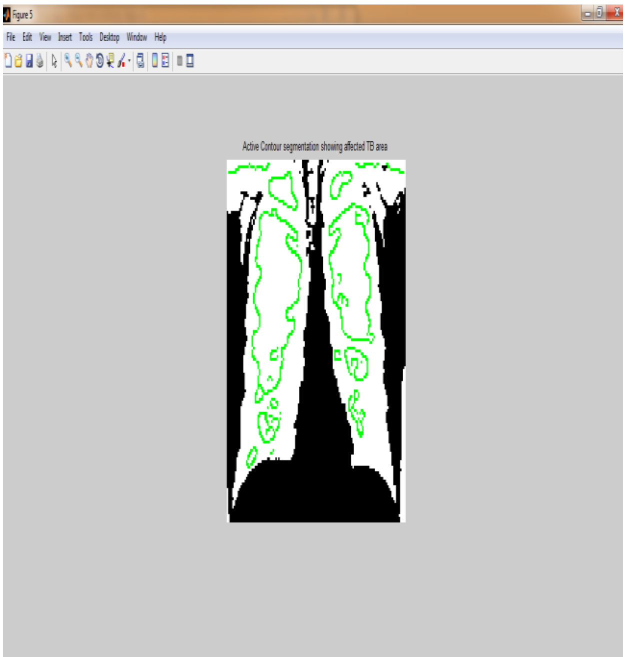

Fig. 6.Active contour lung segmentation

Comparison of both the segmentation and the the performance analysis is shown in fig.7.

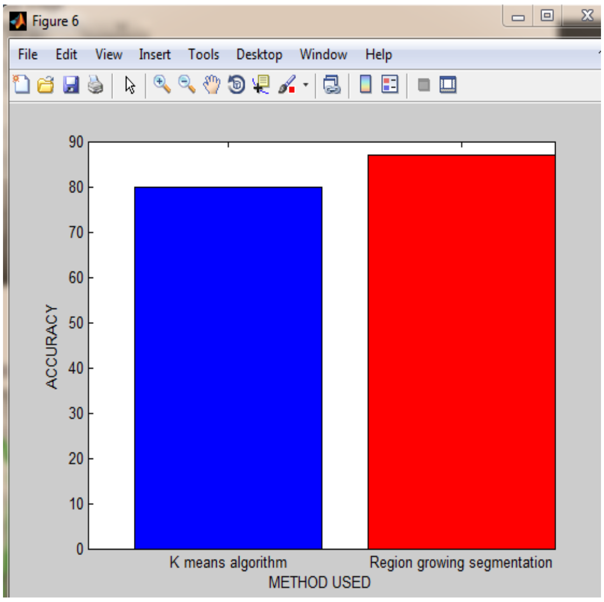

Fig. 7. Comparison of K means and Region growing lung segmentation

V. Conclusion

Here, in this paper we present an automatic detection of tuberculosis. Tuberculosis is the mostly seen in the lung region. So we perform an K means algorithm to segment the lung region from the CXR image which is considered to be the existing work, which is less efficient, we don't get a clear cut of lungs. So we proposed lung segmentation using region growing segmentation algorithm.After performing lung segmentation we perform feature extraction,where various features of the lung region are extracted out. Of these features we consider the best feature mainly the HOG feature. The feature is then fed to a classifier mainly SVM classifier and this SVM classifier is then trained using images from the database and it then perform classification of the CXR image as whether TB is affected or not. Various level of TB affected is also shown according to the input CXR image. The TB affected area is also deermined using region based active contour segmentation. The comparisons between various segmentations are also performed.

Acknowledgment

Anju Mathews thanks P.K.Das Institute of Medical Sciences for providing the CXR database for the successful execution of our work.

.

REFERENCES

[1] B. van Ginneken and B. ter Haar Romeny, "Automatic segmentation of lung fields in chest radiographs," Medical Physics, vol. 27, no. 10, pp.2445–2455, 2000.

[2] B. Van Ginneken, B. ter Haar Romeny, and M. Viergever, "Computer aided diagnosis in chest radiography: a survey," IEEE Trans. Medical Imaging, vol. 20, no. 12, pp. 1228–1241, 2001.

[3] 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 Trans. Medical Imaging, vol. 21, no. 2, pp. 139–149, 2002.

[4] Junji Shiraishi,a Qiang Li, Kenji Suzuki, Roger Engelmann, and Kunio Doi "Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: Localized search method based on anatomical classification"

ISSN: 2278 – 1323

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

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

[7] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in Int. Conf. Comp. Vision Patt. Recog., vol. 1, 2005, pp. 886893.

[8] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. T. H. Romeny, and M. A. Viergever, Active shape model segmentation with optimal features, IEEE Trans. Med. Imag., vol. 21, no. 8, pp.924933, Aug. 2002.

[9] Yonghong Shi, Feihu Qi, Zhong Xue, Liya Chen, Kyoko Ito, Hidenori Matsuo, and Dinggang Shen,Segmenting Lung Fields in Serial Chest Radiographs Using Both Population-Based and Patient-Specific Shape Statistics,IEEE Transactions on medical imaging, vol. 27, no. 4, april 2008

