An Overview on Automatic Tuberculosis Detection

Sethu Parvathi C1, Chithra A S2

1M-Tech Scholar, 2Associate Professor, Dept. of CSE, Lourdes Matha College of Science and Technology, APJ Abdul Kalam Technical University, Kerala, India

Abstract: Tuberculosis is a potentially serious infectious disease caused due to the inhalation of an infectious bacteria- Mycobacterium Tuberculosis. In 2016, an estimated 1.3 million people who were HIV negative died of TB. In addition, there were 374,000 million deaths resulting from TB disease among people who were HIV positive. 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. It is very difficult to diagnose TB by a person’s symptoms on their own. An automatic tuberculosis detection system aids in diagnosing TB especially in remote areas where large number of doctors are not available and also in regions where TB prevalence is high. This paper focuses on a general framework of automatic tuberculosis detection and a review of some of the detection systems proposed.

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 bacterium- 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 TB- latent TB and active TB. Latent TB remains inactive in the body and does not spread from person to person. In active TB, it spreads from person to person and is dreadful if not treated in the right time. There is a small amount of chance that the latent TB can become active TB.

There are several ways of diagnosing TB. A preliminary exam is made by doctor using stethoscope to check for the swelling of lymph nodes. Another one is skin test where in a PPD Tuberculin injection is made in the inside of the forearm. It is checked for red swollen bumps in the arm after 2-3 days which indicates an increased chance of TB. However, it is not 100 percent accurate. Another method is microscopically analysing the sputum samples. In addition to that, blood test known as Interferon Gamma Release Assays( IGRA) can be conducted. It is of high cost and is rarely used.

Out of several methods, chest X- Ray is the cheapest and most reliable method in detecting tuberculosis.

II. LUNGS

Lungs are the primary organs of respiratory system in humans. There are two lungs- left lung and right lung. These are situated in the thoracic cavity of the chest. The right lung is bigger than the left. First, the trachea inhales air into the lungs through its tubular branches called bronchi. The bronchi are divided into smaller branches called bronchioles. The bronchioles end in clusters of microscopic air sacs called alveoli. In the alveoli, oxygen from the air is absorbed into the blood. CO2, a waste product of metabolism, travels from the blood to the alveoli, where it can be exhaled.

The lungs is covered by a thin tissue layer called pleura which act as a lubricant allowing lungs to slip smoothly as they expand contract with each breath. Several diseases affect the lungs and are among them is Tuberculosis. Fig 1 shows the lungs

III. CHEST RADIOGRAPH
Fig 2 shows a normal chest X-Ray. A chest radiograph, also called Chest X-Ray is an imaging test that uses small amounts of radiation to produce pictures of the chest and help diagnose abnormalities or diseases of the airways, blood vessels, bones, heart and lungs. It can also determine if lungs contain fluid or air. Another use of chest X-Ray is to find abnormalities in the size and shape of the heart. Several diseases can be detected based on the abnormalities in X-Ray. It is one of the cheapest and reliable method for detecting tuberculosis. The abnormal findings that can suggest active TB are infiltrates, nodule with poorly defined margins, pleural effusion (presence of a significant amount of fluid within the pleural space).

IV. GENERAL FRAMEWORK OF AUTOMATIC 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 3 shows the flowchart representation of general framework of automatic tuberculosis detection system.

V. RELATED WORKS ON AUTOMATIC TUBERCULOSIS DETECTION SYSTEM

The influence of the medical image in healthcare is constantly growing. Diseases are detected earlier and treatments become more effective. When its usage is expanded beyond the field of diagnostics, entering the arenas of prevention and therapy, it can significantly contribute to lowering costs in healthcare on a global scale. There are several systems that use chest X-Ray in detecting tuberculosis. Chest X-Ray is found to be more reliable. Any manifestations of TB are clearly outlined.

Jaime Melendez, et al [1], proposed a multiple instance learning based approach to CAD of tuberculosis on chest X-Rays. Here, CAD along with supervised learning approach is used. The system includes large databases that contain manually annotated lesions. Manually annotating individual images is a time consuming process and hence multiple instance learning approach is used. The main two stages are Texture Feature Extraction and Pixel Classification. Features collected through texture analysis are called texture feature extraction. It includes normalized horizontal and vertical positions, distance to the lung wall, centre of gravity of both the lungs etc.
Pixel classification is done using k-Nearest Neighbour classifier. After texture classification, a heat map indicating abnormality of each pixel is produced. In this system, manually annotating data for training is a time consuming and tedious task. Diffuse lesions can’t be detected correctly and if new sample data is available, then training process should be repeated again. Disagreement between readers is also a concerning issue. Due to the above limitations, this method is difficult for X-Ray analysis.

Candemir, et al [2], presented a graph cut based lung segmentation method that detects the lungs. The two stages involved in this are average lung shape model calculation and graph cut based segmentation for detection of lung boundaries. Fig 4.1 illustrates the stages of detection system.

Approximate shape model is computed from a set of training images based on its shape similarity. Average of all selected images is the lung model. Second step is to detect the lung boundary. Max flow- min cut theorem is used for segmenting the lung image which results in a global binary segmented image. This method is not accurate and has some robustness because a static shape model is not sufficient to describe the lung region. In addition to that, poor contrast images and images with different lighting conditions do not work well using this system.

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 masks, 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 the complement of an x-ray that lung model mask is a probabilistic lung shape model that is computed using Japanese Society of Radiological Technology(JSRT) dataset.

The log gabor mask is based log gabor wavelets. They attain large spectral information and also maintain maximum spatial localization. These are factors that choose log gabor wavelets over gabor wavelets. The overall lung segmentation mask is the average of the intensity mask, the lung model mask and the log Gabor mask. Segmentation is done with the help of this mask. Following features are extracted for classification-intensity histogram, gradient magnitude histogram, a shape descriptor, a curvature descriptor, angle between the X- axis and largest eigen vector, 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. A linear support vector machine is trained and it classifies an input image into either normal or abnormal. In this method, a static model is created which can cause robustness. The main feature of this approach is that it combine different mask to achieve superior lung segmentation.

Van Ginnekan, et al[4], proposed a hybrid method that combines rule based scheme with a pixel classifier. Here lung fields are segmented using Active Shape Model(ASM).For a given image, n points are described which are known as landmark points. Using set of landmark points, a point distribution model is constructed. Once model is created, the landmark is moved along the direction perpendicular to the contour, i.e. it is moved to n position on either sides and 2n+1 positions are evaluated. The shape model is fitted to points and segmentation is done. Next step is to subdivide the lung field. First, the lung field is divided into upper, middle and lower part. It is done by computing horizontal lines that divide the lung field in three parts of equal area. Then these parts are divided in four sub parts of equal area. It is divided first vertically and then horizontally. Thus 42 regions are defined in the lung field. A multi scale filter bank is used to extract texture features. The filters are 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 classifier is used for classifying regions to normal or abnormal. In this system, segmentation is most accurate at the vertical parts of the rib cage. Failures result when detection areas of diaphragm and lung tops. The segmentation method used in Ginnekan’s work is not suitable in applications that require high accuracy in terms of lung boundary.

An Automated Detection of Early Lung Cancer and Tuberculosis based on X-Ray Image Analysis was put forward by Kim Le, et al[5], where watershed segmentation is used for segmenting the X-Ray image. Watershed segmentation partitions areas into regions.
In this, an input image or a gray scale image can be considered as a topographic surface. Higher intensity denotes peaks and hills and lower intensity denotes valleys. The segmentation process starts from a regional minima, M, the lowest points (valleys) in the area into which the water can flow. With an appropriate distance measure, the area is divided into regions Ω, 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 till the highest points (peaks and hills) can be assigned to any region. The remaining unlabelled points are those of the watershed line. To obtain a thin line, a point is added to the region Ω, even if the distance of that point from the region equals those from other regions. A modified method called water snake approach is used here instead of watershed segmentation which is used to adjust the smoothness of region boundaries. Here gravity measure is introduced to segment a lung X-Ray image. Fig 4.2 and 4.3 represents core parts of lung X-ray before watershed segmentation and after watershed segmentation respectively.

The average and maximal gray levels of the pixels within the scanning window is found. A local threshold is set that has value between the average and the maximal gray levels. After the threshold is set, the number of pixels that have gray level higher than the threshold is calculated. If the counted number falls within a specific range, then early nodule is detected.

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 [4] 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.

VI. CONCLUSION

This paper discusses about the automated tuberculosis detection system, its general framework and a summary of six different works related to it. The purpose of this paper is not to review all related research works in detail but to make you understand the core ideas of these six methods and their advantages and disadvantages [7].

The approach put forward by Jaime Melendez [1], is a multiple instance learning based approach to automatic detection of tuberculosis. This method was time consuming and tedious due to manually annotating data. The representation of data was in the form of heat map and there were disagreement between readers. The main difficulty in any CAD system for X-Ray analysis is segmentation and is one of the major task. The work done by Candemier[2], is based on graph cut based lung segmentation.
A static lung shape model is created. This method has got some robustness and is not accurate. The method does not work well for poor contrast images and images with poor lighting conditions. Stefan Jaegar [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. A linear support vector machine is used for training and classification. Similar to graph cut based segmentation, a static lung model is created in this system that causes robustness. However, it combines different mask to achieve superior lung segmentation. The work done by Kim Le[4] is based on watershed segmentation. An improved approach of watershed segmentation called water snake approach is used to adjust the smoothness of region boundaries. This system also detects small nodules in the lungs which can be an early sign of lung cancer. Van Ginneken[5] proposed a method for finding abnormalities in chest radiographs using local texture analysis. In this system, Active Shape Model(ASM) segmentation is used. Here segmentation shows better result at the vertical parts of the ribcage and poor results near the diaphragm and lung tops. Hence this method is not applicable where accuracy is required in terms of lung boundary. Another work related to automatic detection system was put forward by Binu Joykutty[6] where adaptive thresholding based segmentation is done. This segmentation process leads to different regions of the image to have different threshold. K-NN classifier is used to classify normal/abnormal images. Therefore, computational complexity is less and less time is taken for classification.

The practicality of the works needs to be confirmed although screening tests are done in developing countries and in regions where TB is epidemic. This is an active area of research and has great scope to outperform expert radiologists in the near future.

REFERENCES


