

# Lung Tuberculosis Detection Using X-Ray Images

**Betsy Antony**

*Research Scholar, Department of Computer Science, Christ deemed to be University,  
Hosur Road, Bhavani Nagar, Bengaluru, Karnataka 560029, India.*

*Orcid: 0000-0002-1256-3084*

**Nizar Banu P K**

*Assistant Professor, Department of Computer Science, Christ deemed to be University,  
Hosur Road, Bhavani Nagar, Bengaluru, Karnataka 560029, India.*

## Abstract

This research work is based on the various experiments performed for the detection of lung tuberculosis using various methods like filtering, segmentation, feature extraction and classification. The results obtained from these experiments are discussed in this paper. Lung tuberculosis is a bacterial infection that causes more deaths in the world than any other infectious disease. Two billion people are infected with tuberculosis all around the world. Lung tuberculosis is a disease caused by a bacteria known as *Mycobacterium tuberculosis* or *Tubercle bacillus*. This research work strives to identify methods by which patients, who require second opinion for an already identified result, can save a lot of money. Once we receive X-ray image as an input, pre-processing methods like Gaussian filter, median filter is applied. These filters help to remove unwanted noise and aid to get fine textural features. The output obtained from this is taken as an input and applied to watershed segmentation and gray level segmentation which helps to focus on the lung area of the obtained results. Output from these segmentation methods is fused to get a Region of Interest (ROI). From the ROI, the statistical features like area, major axis, minor axis, eccentricity, mean, kurtosis, skewness and entropy are extracted. Finally, we use KNN, Sequential minimal optimization (SMO), simple linear regression classification methods to detect lung tuberculosis. The results obtained in this paper suggests KNN classifier performs well than the other two classifiers.

**Keywords:** Tuberculosis, X-Ray, graylevel threshold, watershed segmentation, canny edge detection, median filtering.

## INTRODUCTION

With every passing day, almost everything is being automated with the help of machines that can think on their own. Health care is one of the fundamental aspects of human life. Lung tuberculosis is a Bacterial Infection that causes more deaths in the world than any other infectious disease. About 2 billion people are infected with tuberculosis worldwide [1]. Lung

tuberculosis is a disease caused by a bacteria known as *Mycobacterium tuberculosis* or *Tubercle bacillus*. This disease can affect any part of a human body, with the lungs as the primary area of infection. When subjects are infected by tuberculosis bacteria, it kills around 60% of all those who are not treated. Missed diagnoses or delayed diagnoses leads to higher mortality rates. These are aerielly transmitted when a person with active tuberculosis coughs, sneezes or expels air. Symptoms of this disease is coughing blood, weight loss, cough that may last for more than two weeks, fatigue and weakness, fever, night sweats, shortness of breath, chest pain, loss of appetite and chills. Types of tuberculosis(TB) are Active TB Disease, Miliary TB, Latent TB Infection. Active Tuberculosis is when the bacteria spreads rapidly to the different parts of the body. A person with active pulmonary TB disease transmits this to others by airborne transmission of infectious particles spewed into the air. Miliary TB is a rare form of active disease. This type of TB occurs when *Mycobacterium tuberculosis* finds its way into the bloodstream. This bacteria quickly spreads all over the body in tiny nodules and can affect multiple organs at once. Latent TB infection is difficult to be diagnosed as they have no symptoms and their chest x-ray may be normal [2]. If a person is diagnosed with tuberculosis (or without for that matter), it is a human tendency to try and get a second opinion. He can consider the results obtained from this method as a second opinion. This could save a lot of money. X-ray is one of the many diagnostic tools that can be used for detection of lung tuberculosis. This research is focused on developing an application which helps in the detection of lung tuberculosis using Matlab as a research tool. Once the image is loaded, it undergoes several steps such as median filtering, canny filter, segmentation methods such as watershed model and gray level threshold. Later, the output from these tools are fused together. The final step is feature extraction, features such as area, eccentricity, major axis, minor axis, mean, and standard deviation are extracted. The outcome of each step serves as an input for the next step in classification methods. Section II discusses the background, section 3 presents methodology. Section 4 discusses the experimental results carried out in this research. Section 5 concludes this paper.

## BACKGROUND

Adhi Susanto et al [3] have presented an approach to identify lung tuberculosis. Their research work was an attempt to reduce the duration of waiting period for patients who have been waiting for the diagnosis results. To have same textural features in the dataset, image preprocessing methods have been applied. Once an image has been received as input, it is cropped to remove the unwanted black edges in an x-ray. This is then resized in such a way that the lung area in the images are enlarged. Then, the white light intensity emanating from rib bones is subjected to contrasting and then converted to black color. This converts the lung area to black and all the remaining region to white. This is then complemented in such a way that lung region is white and the other areas are black. From the region of interest, statistical features like mean, standard deviation, skewness, kurtosis and entropy are derived. Later, these features are reduced to form principal component analysis (PCA) [3]. The classification method used is minimum distance classifier, which helps to detect if the subject lung has or does not have tuberculosis.

Poomimadevi.CS et al [4] have presented a paper which helps in the detection of tuberculosis. In their research work chest X-Ray (CXR) is taken as the input data. Noise which is present in CXR can be reduced by a filter. But this filtering should not affect the resolution of the image. Therefore restoration filter [3] should be applied. Wiener filter [4] falls under the category of restoration filters and this was used for filtering in this research work. Then, this subject image goes through registration methods like similarity, pattern matching and transformation. To this image, different segmentation methods like watershed segmentation, global thresholding and active contour is applied. Later, a comparative study is made amongst the efficiencies of output with and without registration techniques. Here, images with a combination of registration techniques, watershed and global threshold as batch of segmentation techniques had much accurate results.

Adhi Susanto et al [5] have presented another paper which helps in lung tuberculosis detection. The data set used in this research work is thoracic X-ray images from Dr. Sarjito Hospital, Yogyakarta. Here, Image quality is enhanced using spatial filtering and histogram equalization techniques. Spatial filtering is done for noise reduction and histogram equalization is done for pixel intensity transformation. It also makes intensity level of the dataset uniform. Later, object isolation is done to get the ROI of the image. ROI is considered in different shapes. From these ROIs, features like mean, standard deviation, skewness, kurtosis and entropy are calculated. These five features are reduced to a single feature with the help of PCA. Later minimum distance classification methods have been applied to check if tuberculosis is present in the given dataset.

## METHODS

Methodology followed in this paper to detect lung tuberculosis is represented is figure 1.

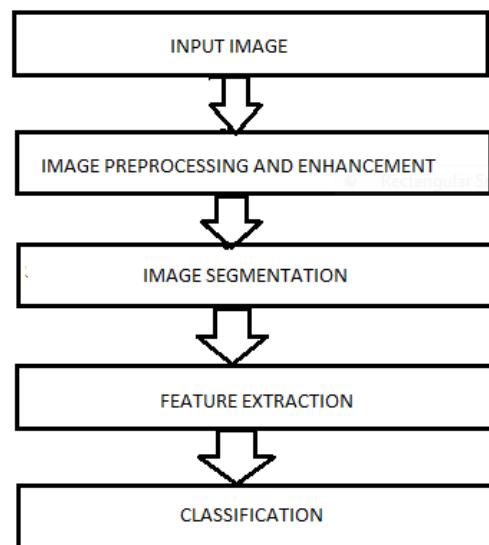


Figure 1: Steps in Image Processing

### Image Preprocessing and Enhancement

The data set we have is not uniform and does not have fine textural features. Therefore, this is the first step in image processing. This is done to get uniformity in the complete dataset. This also enhances the various changes in the images and the different regions have a much higher chance of getting detected. Median filtering [6] is applied, for enhancing the image. This method is to mainly focus on the noise removal operations from the image. 3.1.1

#### Median Filter

In median filtering, the value of an output pixel is determined by the median of the pixel values it is being surrounded by. In median filter the noise gets reduced by square of the number of pixels that we are averaging. The median filter is also a sliding-window spatial filter, which helps with removing noise and pixel transformation. Figure 2 shows the application of median filter on a lung x-ray image.

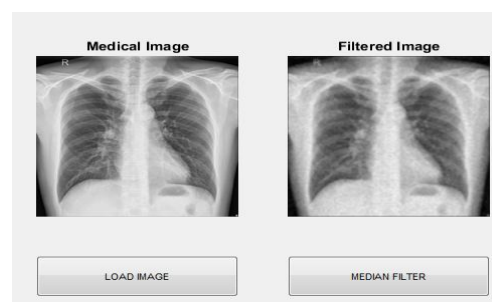
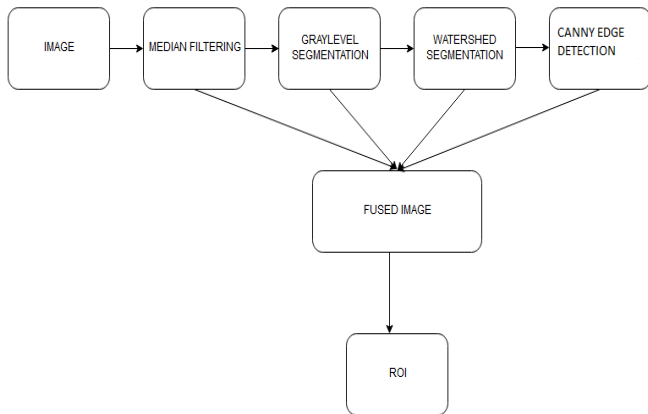


Figure 2: Application of median filter on the lung image

MATLAB is a multipurpose numeric programming language which includes variety of built-in library functions ranging from image processing to higher order numeric calculation. Figure 3 explains the complete procedure done to get the expected output.



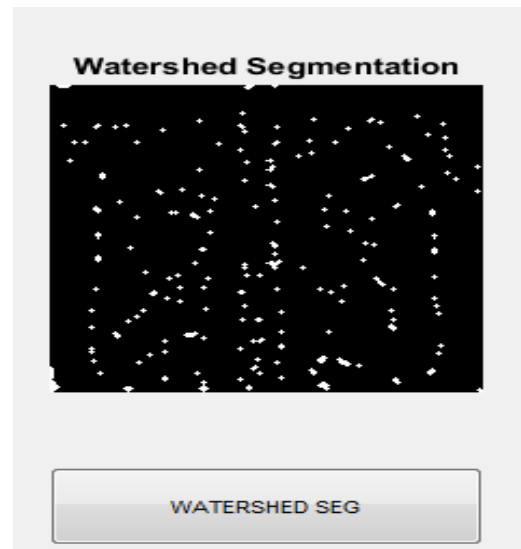
**Figure 3:** Framework of the steps followed

### Image Segmentation

In this section, the images with similar pixel values get grouped together, forming regions. Segmentation is accurate when five qualities are taken care of. 1) Completeness: every pixel should belong to a region. 2) Connectedness: the points of region should be connected with some reason. 3) Disjointedness: There should be some property which differentiates each region. 4) Satisfiability: Pixel of a region must have at least one of the property. 5) Segmentability: two regions should not be merged as one as they have different properties. [7]

### Watershed Segmentation

This segmentation method is applied almost in all medical image processing domains. It is one of the methods which gives accurate results, where grouping is done based on pixel intensities of an image.[8] Watershed is a morphological tool and it is normally used for identifying the outputs rather than using input segmentation techniques. Here, the division is done on the basis of similar attributes. Similarities are separated into different groups. The basic algorithm of watershed segmentation is to transform the gradient of gray level image in a topographic surface. Figure 4 shows application of watershed segmentation for a lung image. [7]



**Figure 4:** Watershed Segmentation when applied on the lung image

### Gray level Threshold Segmentation

Graylevel threshold segmentation converts a gray scale into its binary image[8]. This method is very useful and efficient. This technique basically has a threshold pixel value of 150. Gray level threshold segmentation method is applied for lung images in this paper. The pixel values below 150 is considered as black and pixel values above 150 is considered as white. After thresholding, the image has been segmented into two values 0 and 1 where 0 represents black and 1 represents white. Figure 5 shows application of gray level threshold segmentation for a lung image.



**Figure 5:** Gray-Level Threshold Segmentation applied to the lung image

### Canny Edge Detection

Canny edge detection technique is used to detect the edges of an image. In an image, when there is complete variation in pixel values from its neighbor, then it is considered to be an x-ray image[9]. Canny helps to detect both strong edges and weak edges in a particular image. Through canny filtering structural information is extracted from different objects and therefore reduces the amount of data to be processed [9]. It has been widely applied in various computer vision systems. Canny has found that the requirements for the application of edge detection on diverse vision systems are relatively similar. Figure 6 shows application of canny edge detection for a lung image.

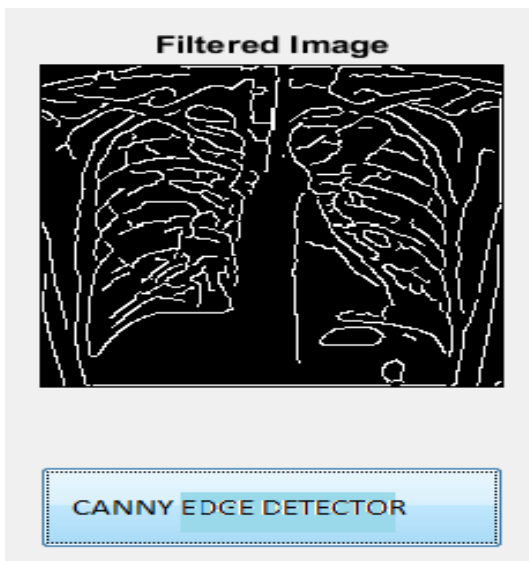


Figure 6: Canny Edge Detection applied to the lung image

### Feature Extraction

The feature extraction technique helps to identify the features in the given image. Regional properties area, major axis, minor axis and eccentricity is calculated. Statistical features like mean, standard deviation, kurtosis and skewness are also calculated.

### Classification

Classification is the final stage in this process. It is the method of categorizing all of the pixels in a digital image into several sets of classes, according to the features selected. K-nearest neighbor (KNN) classification, SMO classification are the different algorithms used in this process.

#### K-Nearest Neighbor(KNN)

K nearest neighbor algorithm helps in identifying the nearest neighbor of an unknown data point. This algorithm works

depending on the value of k. If the value of  $k=n$ , then we can predict n nearest neighbor. In this research work, two classes are present namely, normal class and abnormal class. Using this algorithm, we classify 100 normal images and 100 abnormal images out of 326 normal and 336 abnormal images.

#### Simple Linear Regression

Model statistical-tool used in predicting future values of a target (dependent) variable on the basis of the behavior of a set of explanatory factors (independent variables). A type of regression analysis model, it assumes the target variable is predictable, not chaotic or random. [10]

#### Sequential Minimal Optimization (SMO) Classification

Sequential Minimal Optimization (SMO) is one way to solve the SVM training problem that is quite efficient. SMO uses heuristics to partition the training problem into smaller problems that can be solved analytically. It speeds up the training process[11].

## EXPERIMENTAL RESULTS

The results obtained by applying noise filter, watershed segmentation, gray level threshold segmentation, canny edge detection, feature extraction are discussed in the following sections.

### Dataset

Data used in this research is[12] X-ray images which is taken from National Library of Medicine. There are a total of 662 images available where 326 images are normal and 336 images are abnormal. The details like sex and type of tuberculosis is also available in the given data set. Figure 7 and 8 shows images with TB and without TB.

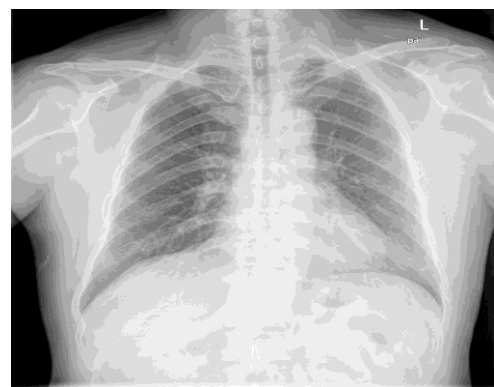


Figure 7: Normal lung image



Figure 8: Abnormal lung image

The output from both these methods along with canny filter is then mixed together to derive a fused output, which gives Region of interest (ROI) from which we find the area, eccentricity, major axis, minor axis, mean, kurtosis, skewness, standard deviation. The expected output image after fusing all methods is shown in figure 10.

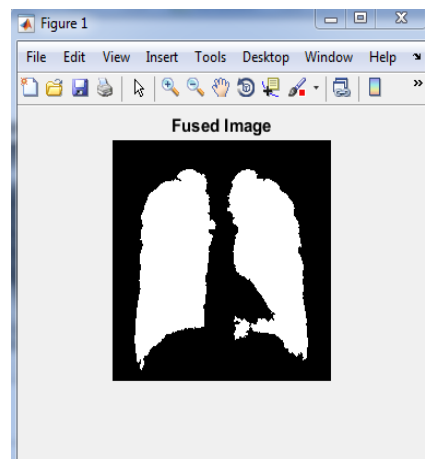


Figure 10: Fused image

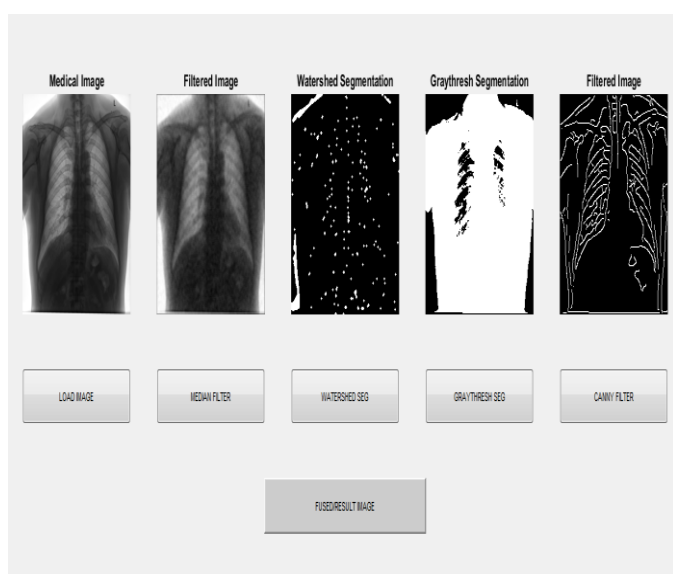


Figure 9: GUI used in detecting lung tuberculosis

Once we receive a thoracic x-ray as input, it is applied to preprocessing method like median filter. Through median filter unwanted noise signal is removed.

The second step is segmentation techniques namely watershed model and gray level threshold. The basic concept of watershed method is to transform the gradient of gray-level image in a topographic surface where the values are interpreted as heights and each local minima in an image is referred as catchments. Likewise In gray-level threshold segmentation, a gray scale image is turned into binary or black and white image by first choosing a gray level threshold value in the input image and then turning every pixel black and white according to whether its gray value is greater than or less than threshold value.

The next step is to detect the edges. Canny is an edge detection technique to extract useful information from different regions and reduce the amount of data to be processed.

Classification is the process of categorizing the input image into either normal class or abnormal class. The classification process is done based on the features selected from the feature extraction. Table 1 shows the obtained accuracy of the classifiers.

Table 1: Classification accuracy

Data set	SIMPLE LINEAR REGRESSION	KNN	SMO CLASSIFICATION
EFFICIENCY	79%	80%	75

## CONCLUSION

A method to detect lung tuberculosis using thoracic x-ray is presented in this research work. To remove unwanted noise from an image, median filtering technique is done at the starting stage. For the next stage we combined two segmentation methods like watershed model and gray level thresholding model, and a fused image is generated which yields a highly accurate result. Features like area, major axis, minor axis, eccentricity, mean, standard deviation, skewness, kurtosis are extracted from ROI of fused image. This is further classified using KNN, SMO and Simple linear regression classifiers. The efficiency of classifiers shows that watershed segmentation and gray level threshold with KNN produces better result with an efficiency of 80% for detecting tuberculosis in lung image. In future, various feature extraction/feature selection methods can be applied for tuberculosis segmentation classification.

**Table 2**

SL No	Classifiers	Accuracy
1	Simple linear regression	79%
2	KNN	80%
3	SMO	75%

From the results obtained it is observed that KNN classifier performs well with 80% accuracy compared to SLR and SMO.

## REFERENCES

- [1] [http://www.who.int/trade/distance\\_learning/gpgh/gpgh3/en/index4.html](http://www.who.int/trade/distance_learning/gpgh/gpgh3/en/index4.html)
- [2] <https://www.nationaljewish.org/conditions/tuberculosis-tb/types>
- [3] Ratnasari Nur Rohmah, Adhi Susanto, Indah Soesanti, Computer Aided Diagnosis for Lung Tuberculosis Identification Based on Thoracic X-ray, IEEE 2013
- [4] Poomimadevi.CS, Helen Sulochana.C, Automatic detection of pulmonary tuberculosis using image processing techniques, IEEE WiSPNET 2016
- [5] Ratnasari Nur Rohmah, Adhi Susanto, Indah Soesanti, Lung Tuberculosis Identification Based on Statistical Feature of Thoracic X-ray, IEEE 2013
- [6] [https://en.wikipedia.org/wiki/Median\\_filter](https://en.wikipedia.org/wiki/Median_filter)
- [7] R.C Gonzalez, Digital Image Processing. Pearson, 2014, pp.689-795.
- [8] Rohini Paul Joseph, M Manikandan, Brain tumor MRI image segmentation and detection in image processing, International Journal of Research in Engineering and Technology, volume 3, special issue 01, March 2014.
- [9] [https://en.wikipedia.org/wiki/Canny\\_edge\\_detector](https://en.wikipedia.org/wiki/Canny_edge_detector)
- [10] <http://www.businessdictionary.com/definition/classical-linear-regression-CLR.html>
- [11] <https://stats.stackexchange.com/questions/130293/svm-and-smo-main-differences>
- [12] <https://ceb.nlm.nih.gov/repositories/tuberculosis-chest-x-ray-image-data-sets/>