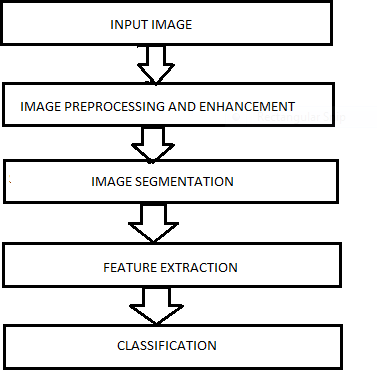
**i) Introduction:**

* 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 aerially 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.

**ii) Methods:**

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

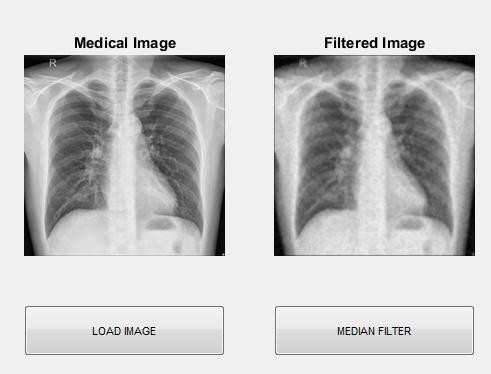


# Image Preprocessing and Enhancement

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

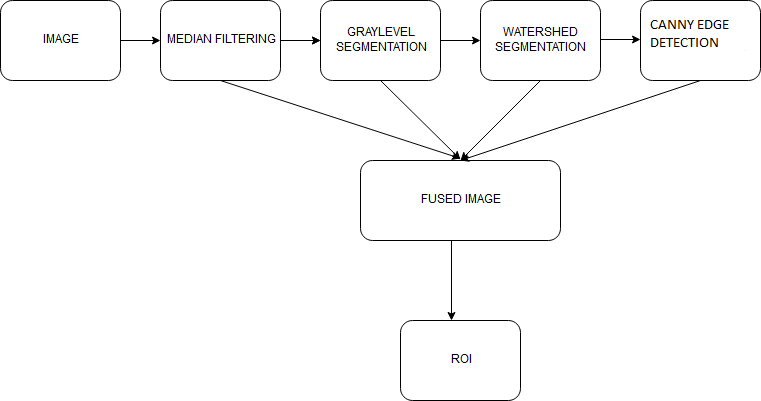
**Median Filter**

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.

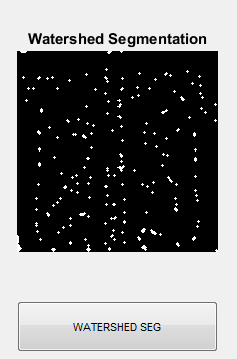
  
  
 ***Figure 3****: Framework of the steps followed*

## 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]

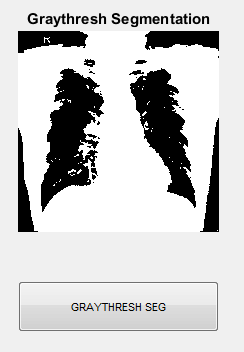
# Watershed Segmentation

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*

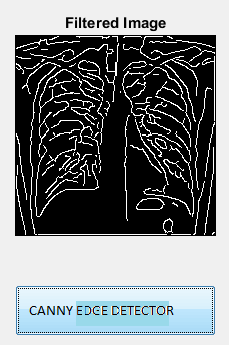
## Gray level Threshold Segmentation

Gray level threshold segmentation converts a gray scale into its binary image [8]. The is 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.

epending 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 and100 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].

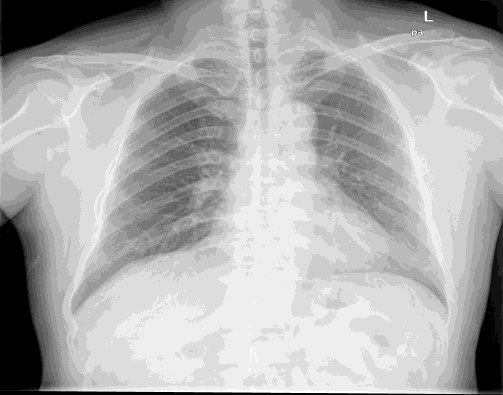
# iii) Results:

The results obtained by applying noise filter, watershed segmentation, gray level threshold segmentation, canny edge detection, feature extraction is discussed.

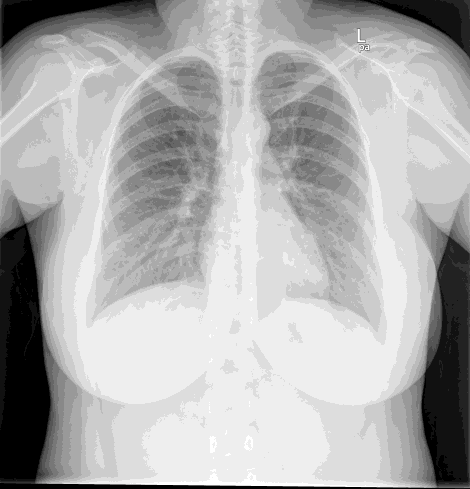
# iv) Discussion/Conclusion

## Dataset

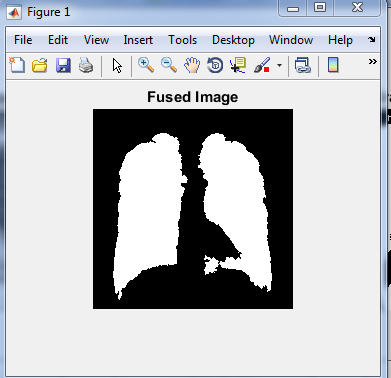
Data used in this research is [12] X-ray images which is taken from National Library of Medicine. There are a total of 662 images available where 326 images are normal and 336 images are abnormal. The details like sex and type of tuberculosis are also available in the given data set. Figure 7 and 8 shows 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*

# 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.

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