Analysis of Tuberculosis in chest using SVM classifier

S.Sivaranjani
M.E-APPLIED ELECTRONICS
J.J College of Engineering and Technology
Trichy,Tamilnadu,India

Abstract— Tuberculosis is a major health threat in many regions of the world after HIV, while diagnosing tuberculosis still remains a challenge. It is estimated that nine million new cases appear every year. Mortality rates of patients with undiagnosed TB are high. Modern diagnostic techniques are often too slow and too expensive. In order to reduce the burden of the disease, an automated approach for detecting TB on conventional posteroanterior chest radiographs is developed. The chest radiograph (CXR) combines all thoracic analysis and furnish a high crop, given the low price and single origin. In the proposed method, wiener filter is used to remove noise and the lung region is extracted using graph cut segmentation. Graph cut segmentation method is preferred because segmentation of medical images is very difficult. For this lung region, texture and shape features are computed which enables the CXRs to be classified using multi class SVM. The extracted features are local binary patterns (LBP), histogram of oriented gradients (HOG) and tamura texture descriptor. Multiclass SVM classify the chest region into normal, beginning stage, moderate stage and severe stage.

Keywords-Tuberculosis, chest radiographs (CXRs)

I. INTRODUCTION

Tuberculosis is second major health threat in many regions of the world. About one-third of the world's population has latent TB, which means people have been infected by TB. Tuberculosis (TB) is caused by bacteria (Mycobacterium tuberculosis) that most often affect the lungs. Tuberculosis is curable and preventable. Over 95% of cases and deaths are in developing countries. People who are infected with HIV are 26 to 31 times more likely to become sick with TB. Risk of active TB is also greater in persons suffering from other conditions that impair the immune system. Over half a million children (0-14 years) fell ill with TB, and 80 000 HIV-negative children died from the disease in 2013. Tobacco use greatly increases the risk of TB disease and death. More than 20% of TB cases worldwide are attributable to smoking.

Common symptoms of active lung TB are cough with sputum and blood at times, chest pains, weakness, weight loss, fever and night sweats. Many countries still rely on a long-used method called sputum smear microscopy to diagnose TB. Trained laboratory technicians look at sputum samples under a microscope to see if TB bacteria are present. With three such tests, diagnosis can be made within a day, but this test does not detect numerous cases of less infectious forms of TB. Diagnosing MDR-TB and HIV-associated TB can be more complex. In addition several skin tests based on immune response are available for determining tuberculosis. But skin tests are not always reliable. The latest development for detection are molecular diagnostic tests that are fast and accurate, and that are highly sensitive and specific. The initial work on computer aided diagnosis in chest radiology, there are still no systems that can accurately read chest radiographs. Automated nodule detection is becoming one of the more mature applications of decision support for CXR and CT. The segmentation of the lung field is a typical task that any CAD system needs to support for proper evaluation of CXRs.

II. EXISTING WORK

In this paper diagnosis of Tuberculosis is done by image processing techniques and classification of lungs is done by multi class support vector machine. Modern diagnostic methods take more time to produce the diagnosis results. In order to reduce the burden of the disease, an automated approach for detecting TB on conventional posteroanterior chest radiographs is developed. Clavicle region is a notoriously difficult region for TB detection because the clavicles can obscure manifestations of TB in the apex of the lung. A cavity in the
upper lung zones is a strong indicator that TB had developed into highly infectious state. In existing work discrimination of lungs is done by binary support vector machine and diagnose whether the lung is normal or abnormal. First lung region is extracted by means of graph cut segmentation. Then shape and texture features are computed from the segmented lung region. Features that are computed are local binary pattern, histogram of oriented gradients and tamura features. Finally classification is done by binary SVM classifier whether it is normal or abnormal. Class labels of binary SVM classifier takes two values +1, -1. Abnormal CXRs have positive value and normal CXRs have negative value. Accuracy of ROC curve is about 79%.

III. PROPOSED WORK

In the proposed work multi class SVM is used to classify the chest radiographs (CXR) as normal stage, beginning stage, moderate stage and severe stage of tuberculosis in lungs. CT images (computed tomography) are given as input. Wiener filter is used to filter noise. Filtered image is segmented by graph cut segmentation. Graph cut segmentation method is preferred because segmentation of medical images are very difficult. From the segmented lung image three shape and texture features are extracted. These feature are local binary pattern, histogram of oriented gradient and tamura feature. Finally classification is done. The accuracy of ROC curve is about 83%.

A. Preprocessing

Preprocessing images commonly involves removing low frequency back ground noise normalizing the intensity particle images, removing reflections and masking portions of images. Image processing is the technique of enhancing data images prior to computational processing. In this section wiener filter is applied to the input CT images of lungs. In image processing, the Wiener filter is a filter used to produce an estimate of a desired or target random process by linear time-invariant filtering of an observed noisy process, assuming known stationary image and noise spectra, and additive noise. The Wiener filter minimizes the mean square error between the estimated random process and the desired process. It is used remove Gaussian noise and white noise.

B. Graph cut segmentation

Various methods have been applied to detect the lung boundary on x-ray chest images. The methods are rule based methods, hybrid method, deformable model based methods and pixel based methods. The Lung segmentation is an optimization problem that takes properties of lung boundaries, regions and shape into account. In general, segmentation in medical images has to cope with poor contrast, acquisition noise due to hardware constraints and anatomical shape variations. As applied in the field of computer vision, graph cuts can be employed to efficiently solve a wide variety of low-level computer vision problems such as image smoothing, the stereo correspondence problem. Therefore lung model represents the average lung shape model of selected training masks. These masks are selected according to similar shape. Then vertical and horizontal intensity projections of the histogram equalized images are computed. Graph cut algorithm which models the segmentation problem with an objective function. The objective function is described in terms of boundary, region and shape model properties. The algorithm computes a binary segmentation by minimizing the objective function. Based on the system criteria, objective function is defined in terms of boundary region and shape model properties of the pixels as follows

\[ E(f) = E_d(f) + E_s(f) + E_m(f) \]

Where \( E_d(f) \) = data term \( E_s(f) \) = smoothness term \( E_m(f) \) = lung model term
C. Feature Extraction

To describe normal, beginning stage, moderate stage and severe stage patterns in the segmented lung field. The extracted features are local binary pattern, histogram of oriented gradients and tamura texture descriptor. Local binary patterns (LBP) are a type of feature used for classification in computer vision. LBP is the particular case of the Texture Spectrum model proposed. The LBP operator assigned a label to every pixel of a gray level image. The label mapping to a pixel is affected by the relationship between this pixel and its eight neighbors of the pixel.

Histogram of Oriented Gradients (HOG) is feature descriptors used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. The first step of calculation is the computation of the gradient values. The second step of calculation involves creating the cell histograms.

The tamura descriptor is motivated by the human visual perception. The descriptor comprises a set of six features and commonly three features are extracted such as contrast, directionality and coarseness. Coarseness relates to distances of notable spatial variations of grey level. Degree of directionality is measured using the frequency distribution of oriented local edges against their directional angles. Contrast measures the gray levels

D. Classification

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns used for classification and regression analysis. CT abnormality detection is done by multiclass SVM. Each SVM in the multi-class classifier is trained separately to achieve its best classification performance by choosing proper features before they are aggregated. In this approach multiclass SVM used for classifying the chest region into Normal, Beginning stage, Moderate stage and severe stage.

IV. CONCLUSION

Thus an automated analysis that screens CXRs for manifestations of TB. When given a CXR as input, system first segments the lung region using an optimization method based on graph cut. Each SVM in the multi-class classifier is trained separately to achieve its best classification performance by choosing proper features before they are aggregated. The results of the experiments indicate that this method has much faster training and testing times than the widely used multi-class SVM methods like one-against-one and one-against-all while keeping comparable recognition rates. This approach is an optimization method that combines data with illustrate lung atlas models and classifies the Normal, Beginning stage, Moderate stage and Severe stage from the given input and provides best classification performance.

V. REFERENCES


