International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 4, No. 1, Mar. - May, 2019
Pages 6-13 | Cosmos Impact Factor (Germany): 5.195
Received: 02.03.2019 Published : 21.03.2019

AUTOMATIC IDENTIFICATION OF TUBERCULOSIS AFFECT IN LUNGS IMAGES USING SUPPORT VECTOR MACHINE

[1]P.Priyadharshini [2]R.S.Priyanka [3]G.Radhika [4]S.Kalaivani
Department of Electronics and Communication Engineering,
Rajalakshmi Institute of Technology, Tamil Nadu, India

ABSTRACT

Tuberculosis (TB) is very dangerous and rapidly spread disease in the world.Opportunistic infections in immune compromised HIV/AIDS patients and multi-drug-resistant bacterial strains have exacerbated the problem, while diagnosing tuberculosis still remains a challenge. When left undiagnosed and thus untreated, mortality rates of patients with tuberculosis are high.In the investigating cases for suspected tuberculosis (TB), chest radiography is not only the key techniques of diagnosis based on the medical imaging but also the diagnostic radiology. So, Computer aided diagnosis (CAD) has been popular and many researchers are interested in this research areas and different approaches have been proposed for the TB detection and lung disease classification. In this project the automatic identification of tuberculosis affect lungs images are proposed. In an effort to reduce the burden of the disease, this paper presents our automated approach for detecting tuberculosis in SVM method. We first extract the lung region using a noise in pre-processing method. For this lung region, we compute into threshold method. Following that feature extraction and x-rays to be classified as normal or abnormal using a SVM classifier. The main aim of the method is to develop a CAD (Computer Aided Diagnosis) system for finding the lung tumour using the lung CT images and classify the tumour as beginning.

Keywords : computer-aided detection ,dataset ,segmentation, feature extraction and classification(SVM).

INTRODUCTION

Tuberculosis is an infectious disease caused by the bacillus Mycobacterium tuberculosis which affects mainly the lungs. In 2017 over 8.7million people fell ill with TB and 1.4million died. Large number of Patients with TB need to be X-ray is an inexpensive way to screen for the presence of TB the interpretation of x-ray is subject to human error.

Tuberculosis (TB) is an infectious disease caused by the bacterium Mycobacterium tuberculosis (MTB). Tuberculosis generally affects the lungs, but can also affect other parts of the body. Most infections do not have symptoms, known as latent tuberculosis. About 10% of latent infections progress

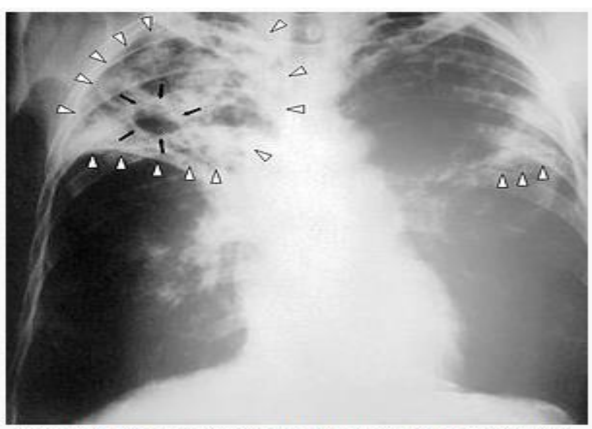
International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 4, No. 1, Mar. - May, 2019
Pages 6-13 | Cosmos Impact Factor (Germany): 5.195
Received: 02.03.2019 Published : 21.03.2019

to active disease which, if left untreated, kills about half of those people who are infected. The classic symptoms of active TB are a chronic cough with blood-containing sputum, fever, night sweats, and weight loss.

Tuberculosis is spread through the air when people who have active TB in their lungs cough, spit, speak, or sneeze. People with latent TB do not spread the disease. Active infection occurs more often in people with HIV/AIDS and in those who smoke. Diagnosis of active TB is based on chest X-rays, as well as microscopic examination and culture of body fluids. Diagnosis of latent TB relies on the tuberculin skin test (TST) or blood tests.

Prevention of TB involves screening those at high risk, early detection and treatment of cases, and vaccination with the bacillus Calmette-Guérin vaccine. Treatment requires the use of multiple antibiotics over a long period of time. Antibiotic resistance is a growing problem with increasing rates of multiple drug-resistant tuberculosis (MDR-TB).

One-third of the world's population is thought to be infected with TB. New infections occur in about 1% of the population each year. In 2014, there were 9.6 million cases of active TB which resulted in 1.5 million deaths. More than 95% of deaths occurred in developing countries. The number of new cases each year has decreased since 2000. About 80% of people in many Asian and African countries test positive while 5–10% of people in the United States population tests positive by the tuberculin test.



Chest X-ray of a person with advanced tuberculosis: Infection in both lungs is marked by white arrow-heads, and the formation of a cavity is marked by black arrows.

Figure- Lung affected with tuberculosis

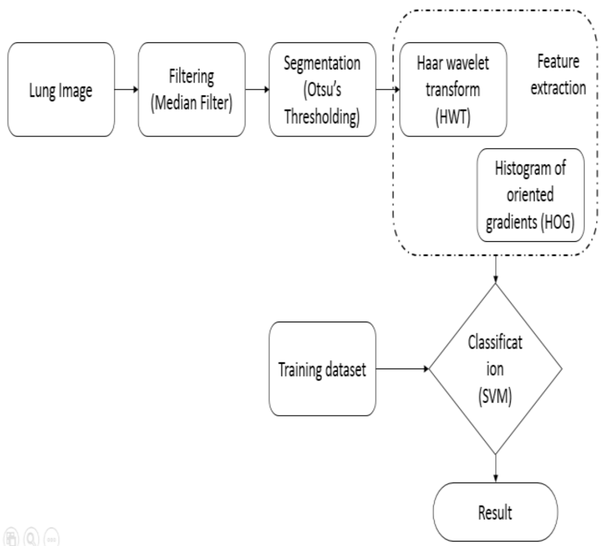
The majority of all cases are caused by tobacco smoking. Exposure to asbestos, radon, uranium and arsenic are other factors for lung cancer. Lung cancer is a deadly disease and has chances to spread to other parts of the body, e.g. the brain, liver, bone and bone marrow. Mass screening of a large population is a time consuming and tedious task, for this reason there is developing tool is computer–aided diagnostic system(CAD) that detect automatically. The early detection and diagnosis of nodules in CT image are among the most challenging clinical tasks performed by radiologists.[1]

The complexity for finding the lung nodules in radiographs are as follows:

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 4, No. 1, Mar. - May, 2019
Pages 6-13 | Cosmos Impact Factor (Germany): 5.195
Received: 02.03.2019 Published : 21.03.2019

1. A nodule diameter may be differed from a few millimeters
2. Nodules vary widely in density.
3. As nodules can be found anywhere in the lung region, they can be hidden by ribs and structures below the diaphragm, resulting in a large variation of contrast to the background.
4. To overcome these difficulties, the author proposed a Computer Aided Diagnosis (CAD) system for detection of lung nodules .The lung tumour prediction[2].

FLOW CHART



Explanation

MEDIAN FILTER

The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction is a typical pre-processing step to improve the results of later processing (for example, edge detection on an image). Median filtering is very widely used in digital image processing because, under certain conditions, it preserves edges while removing noise.

SEGMENTATION:

In computer vision image segmentation is the process of partitioning digital images into multiple segments (set of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 4, No. 1, Mar. - May, 2019
Pages 6-13 | Cosmos Impact Factor (Germany): 5.195
Received: 02.03.2019 Published : 21.03.2019

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contour extracted from the image. Each of the pixels in a region are similar with respect to some characteristic or computed property, such as colour, intensity, or texture .

OTSUS THRESHOLDING:

In computer vision and image processing, Otsu's method, named after Nobuyuki Otsu is used to automatically perform clustering-based image thresholding or the reduction of a graylevel image to a binary image. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal, or equivalently (because the sum of pairwise squared distances is constant), so that their inter-class variance is maximal. Consequently, Otsu's method is roughly a one-dimensional, discrete analog of Fisher's Discriminant Analysis.
The extension of the original method to multi-level thresholding is referred to as the multi Otsu method.
Method

In Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t)$$

Weights omega _{0,1} are the probabilities of the two classes separated by a threshold t and sigma _{0,1}^{2} are variances of these two classes.

The class probability omega _{0,1}(t) is computed from the L histograms:

$$\omega_0(t) = \sum_{i=0}^{t-1} p(i)$$

$$\omega_1(t) = \sum_{i=t}^{L-1} p(i)$$

Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance:

$$\begin{aligned}\sigma_b^2(t) &= \sigma^2 - \sigma_w^2(t) = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \\ &= \omega_0(t)\omega_1(t)[\mu_0(t) - \mu_1(t)]^2\end{aligned}$$

which is expressed in terms of class probabilities omega and class means mu . while the class mean mu _{0,1,T}(t) is:

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 4, No. 1, Mar. - May, 2019
Pages 6-13 | Cosmos Impact Factor (Germany): 5.195
Received: 02.03.2019 Published : 21.03.2019

$$\mu_0(t) = \sum_{i=0}^{t-1} i\frac{p(i)}{\omega_0}$$

$$\mu_1(t) = \sum_{i=t}^{L-1} i\frac{p(i)}{\omega_1}$$

$$\mu_T = \sum_{i=0}^{L-1} ip(i)$$

The following relations can be easily verified:

$$\omega_0\mu_0 + \omega_1\mu_1 = \mu_T$$
$$\omega_0 + \omega_1 = 1$$

The class probabilities and class means can be computed iteratively. This idea yields an effective algorithm.

HAAR WAVELET TRANSFORMATION:

Haar function are simplest wavelets these forms are used in many methods of discrete image transforms and processing. The image transform theory is a well known area characterized by a precise mathematical background, but in many cases some transforms have particular properties which are not still investigated.

Analysis of the Two-Dimensional HWT

You can see why the wavelet transformation is well-suited for image compression. The two-dimensional HWT of the image has most of the energy conserved in the upper left-hand corner of the transform - the remaining three-quarters of the HWT consists primarily of values that are zero or near zero. The transformation is local as well - it turns out any element of the HWT is constructed from only four elements of the original input image. If we look at the HWT as a block matrix product, we can gain further insight about the transformation.

Suppose that the input image is square so we will drop the subscripts that indicate the dimension of the HWT matrix. If we use H to denote the top block of the HWT matrix and G to denote the bottom block of the HWT, we can express the transformation as:

$$B = WAW^T = \begin{bmatrix} H \\ G \end{bmatrix} A \begin{bmatrix} H \\ G \end{bmatrix}^T = \begin{bmatrix} H \\ G \end{bmatrix} A \begin{bmatrix} H^T \\ G^T \end{bmatrix} = \begin{bmatrix} HA \\ GA \end{bmatrix} \begin{bmatrix} H^T \\ G^T \end{bmatrix} = \begin{bmatrix} HAH^T & HAG^T \\ GAH^T & GAG^T \end{bmatrix}$$

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 4, No. 1, Mar. - May, 2019
Pages 6-13 | Cosmos Impact Factor (Germany): 5.195
Received: 02.03.2019 Published : 21.03.2019

We now see why there are four blocks in the wavelet transform. Let's look at each block individually. Note that the matrix H is constructed from the lowpass Haar filter and computes weighted averages while G computes weighted differences.

The upper left-hand block is HAHT - HA averages columns of A and the rows of this product are averaged by multiplication with HT. Thus the upper left-hand corner is an approximation of the entire image. In fact, it can be shown that elements in the upper left-hand corner of the HWT can be constructed by computing weighted averages of each 2 x 2 block of the input matrix. Mathematically, the mapping is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow 2\cdot(a+b+c+d)/4$$

HISTOGRAM OF ORIENTED GRADIENTS:

The histogram of oriented gradients (HOG) is a feature descriptor 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. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

Navneet Dalal and Bill Triggs, researchers for the French National Institute for Research in Computer Science and Automation (INRIA), first described HOG descriptors at the 2005 Conference on Computer Vision and Pattern Recognition (CVPR). In this work they focused on pedestrian detection in static images, although since then they expanded their tests to include human detection in videos, as well as to a variety of common animals and vehicles in static imagery.

FEATURE EXTRACTION*:* This section explains the various features extracted from the feature extraction process. The features, such as LGP, length, area, density and various histogram features, such as color, mean and variance, are the important features extracted from the segmented. In image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction.

SVM classifier

Support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a

International Research Journal in Global Engineering and Sciences. (IRJGES)
ISSN : 2456-172X | Vol. 4, No. 1, Mar. - May, 2019
Pages 6-13 | Cosmos Impact Factor (Germany): 5.195
Received: 02.03.2019 Published : 21.03.2019

representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The clustering algorithm which provides an improvement to the support vector machines is called support vector clustering and is often used in industrial applications either when data are not labeled or when only some data are labeled as a preprocessing for a classification pass.

Conclusion

Pulmonary tuberculosis (TB) persists as a great public health problem in Korea. Increases in the overall age of the population and the rise of drug-resistant TB have reinforced the need for rapid diagnostic improvements and new modalities to detect TB and drug-resistant TB, as well as to improve TB control. An automatic method is presented to detect abnormalities in frontal chest radiographs which are aggregated into an overall abnormality score. The method is aimed at finding abnormal signs of a diffuse textural nature, such as they are encountered in mass chest screening against tuberculosis (TB). The scheme starts with filtering the chest radiographs using median filter then the features of the filtered image will be extracted using Haar wavelet transform (HWT) and histogram of oriented gradients then the automatic segmentation of the lung fields will be done by using Otsu's thresholding algorithm. Finally the extracted futures are compared with the trained datasets using support vector machine and the result will be shown whether then tuberculosis is affected or not. All the above process completed and lung images are successfully classified whether the lung image was affected or not by using support vector machine

VI. References

[1]Stefan Jaeger, Alexandros Karargyris, Serna Candemir, Les Folio, Jenifer Siegelman, Fiona Caliaghan,Zhiyun Xue, Kannappan Palaniappan, Rahul K. Singh, Sameer Antani, George Thoma, Yi-Xiang Wang,Pu-Xuan Lu, and Clement J. McDonald, "Automatic Tuberculosis Screening Using Chest Radiographs", IEEE transactions on medical imaging, vol. 33, no. 2, February 2014

[2]Wai Yan Nyein Naing, Zaw Z. Htike, "Advances in automatic tuberculosis detection in chest x-ray images", signal & image 802 processing an international journal (sipij) vol.5, no.6, December 2014.

[3]Rui Shen, , Irene Cheng, , and Anup 8asu, Senior Member, "A Hybrid Knowledge-Guided Detection Technique for Screening of Infectious Pulmonary

Tuberculosis From Chest Radiographs". IEEE transactions on biomedical engineering, vol. 57, no.I I, November 2010.

[4]Tao Xu, Irene Cheng, Senior Member IEEE, and Mrinal Mandai, Senior Member, IEEE, "Automated Cavity Detection of Infectious Pulmonary Tuberculosis in Chest Radiographs", 33rd Annual international Conference of the IEEE EMBS Boston, Massachusells USA, August 30 - September 3, 2011.

[5]A.M. Khan, Ravi. S ," Image Segmentation Methods: Comparative Study", International Journal of Soft Computing and Engineering (USCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013.

[6]Serna Candemir, Stefan Jaeger, Kannappan Palaniappan, Sameer Antani, George Thoma, and Clement J. McDonald," LungSegmentation in Chest Radiographs Using Anatomical Atlases With Nonrigid Registration", iEEE transactions on medical imaging, vol. 33, no. 2, february 2014.

[7]Hrudya Dasl ,Ajay Nath2 ,"An Efficient Detection of Tuberculosis from Chest X-rays ", international Journal of Advance Research in Computer Science and Management Studies", Volume 3, Issue 5, May 2015 pg. 149-154.

[8]P. Maduskar, H. Laurens, R. Philipsen, and B. Ginneken, Automated localization of costophrenic recesses and costophrenic angle measurement on frontal chest radiographs, inProc. SPIE, 2013, vol. 8670.

[9]A.Gururajan, H. Sari-Sarraf, and E. Hequet, Interactive texture segmentation via IT-SNAPS, inIEEE Southwest Symp. Image Anal. Interpret., 2010.

[10]L. Hogeweg, C. Snchez, P. A. Jong, P. Maduskar, and B. Ginneken,Clavicle segmentation in chest radiographs,Med. Image Anal., vol.16, no. 8, pp. 1490-1502, 2012.

[11]M. Rahman, S. Antani, D. Fushman, and G. Thoma, “Biomedical image retrieval using multimodal context and concept feature spaces,”Med. Content-Based Retrieval Clin. Decision Support, vol. 7075, pp.24-35, 2012.

[12]S. Candemir and Y. S. Akgul, Statistical significance based graph cut regularization for medical image segmentation,TurkishJ.Electr.Eng.Comput. Sci., vol. 19, no. 6, pp. 957-972, 2011.

[13]C. Liu, J. Yuen, and A. Torralba, “SIFTflow: Dense correspondence across different scenes and its applications.,IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, 2011.

[14]A.Dawoud, Lung segmentation in chest radiographs by fusing shape information in iterative thresholding,IET Comput. Vis., vol. 5, no. 3, pp. 185–190, 2011.

[15]S.Candemir, S.Jaeger, K.Palaniappan, S.Antani, and G.Thomas, Graph-cut based automatic lungboundary detection in chest radiographs, in IEEE Healthcare Technol. Conf.: Translat. Eng. Health Med., 2012, pp. 31-34.