ISSN (e): 2250 – 3005 || Volume, 08 || Issue, 7|| July – 2018 ||
International Journal of Computational Engineering Research (IJCER)

Detection of Tuberculosis Using Novel Multiple Instance Learning Based Approach on Chest X-Rays

Mr.Onkar A. Badadare[1], Prof. (Dr.) R.N.Patil[2]
[1, 2] Department of Electronics Engineering, D.K.T.E's Textile and Engineering Institute, Ichalkaranji, Maharashtra state, India
Correspondence Author: Mr.Onkar A. Badadare

ABSTRACT:
Tuberculosis is dangerous and widely spread disease in the world. Manual detection of Tuberculosis using chest X-rays is tedious task and require specialized person which is not always available. So, it needs of hourat all places a Computer Aided Detection (CAD) system which requires less human interaction. Using Multiple Instance Learning (MIL) this is applied to CAD system for detection of Tuberculosis.miSVM is one method of MIL which has several drawbacks. To overcome miSVM drawback probability estimation is used with miSVM. In this paper, we compare SVM, miSVM and miSVM+PEDD method based on image scoring and CPU time.
Keywords: Chest X-rays, Computer Aided Detection (CAD), feature extraction, Multiple Instance Learning (MIL), region growing segmentation, Support Vector Machine(SVM),Tuberculosis (TB)

Date of Submission: 04-07-2018 Date of acceptance: 21-07-2018

I.INTRODUCTION

Tuberculosis is caused by Mycobacterium Tuberculosis and spread from infectious person through air. Symptoms of this disease include cough, bleeding of lung, fever, loss of weight and weakness. Due to high infectivity early detection and treatment are best prevention for TB disease. In this paper, we have presented automatic approach for detection and assigning image score to each chest X-ray image for detecting Tuberculosis. Image score range starts from 0 to 100.Image score below 50 shows normal X-ray image(No TB) and above 50 shows abnormal image(TB detected).CPU time is time required for CAD system to calculate image score. Three classifier areused for detecting abnormality and calculating image score in chest X-rays.

II.METHODOLOGY

A.Training database

Zambia database is used for training as well as for testing. It consists of 50 normal and 50 abnormal images.24 images used in test image set called as Zambia test set. Zambia database is first goes to training. Training involves loop wise pre-processing, region growing segmentation, calculating LBP feature and feature extraction of each image. After training trained features are calculated by CAD system.

B. Test image and pre-processing

Test image is taken from test set and this image undergoes for pre-processing. Pre-processing involves noise removal using wiener filter and contrast enhancement of test image.

C. Region growing segmentation and cavity detection

Region growing segmentation is a simple region based segmentation method.It is also classified as a pixel based image segmentation since it involves selection of seed point.The first step in region growing segmentation is to select a set of seed points.Seed point selection is based on different criteria (pixels in a certain grayscale range, pixels evenly spaced on a grid, etc.).The initial region starts as exact location of this seeds. The region grows from these seed points to adjacent points depending on pixel intensity, grayscale, texture. We have used eight connected neighborhoods for the pixels adjacent relationship. At last cavity is detected. This cavity also called as Region of Interest (ROI).

D. Local Binary Pattern (LBP) and feature extraction

LBP features are calculated. For each pixel in a cell, the pixel compares to each of its eight neighbors (on its left-top, left-middle, left-bottom, right-top, etc.).By following the pixels along a circle, i.e. clockwise or counter-clockwise. Where the center pixel's value is greater than the neighbor, write "1". Otherwise, write "0".

This gives an eight digit binary number (which is usually converted to decimal for convenience). Then histogram iscomputed over the cell. The feature vectors are calculated by normalizing and concatenating histogram.463 LBP features for each image are extracted during feature extraction process.

E. Classification Methods

1. Support Vector Machine (SVM) classifier

The classification of obtained feature vectors is carried out using SVM classifier. The main goal of SVM is to design hyperplane that classify all training vectors in two classes with maximum margin from both classes.

2. miSVM classifier

miSVM is nothing but Multiple Instance Learning (MIL) with SVM.MIL is generalization of supervised classification that associates training class label with set of pattern (bags) instead of individual pattern. In binary classification; a bag is labeled positive if at least one of its instances is positive and bag is labeled negative if all instances in it are negative.

Given instance label $y_{i,}$ $y_i \in \{-1, 1\}$, i=1, 2,…….N

N-Number of training instances xi & bag labelsYI,YI $\in\{-1, 1\}$ &I=1, 2……., M

M-Number of training bags BI, BI = $\{x_i: i \in I\}$

miSVM formulation can be written as,

min min $\frac{1}{2}||w||^2 + C\sum_i \xi_i$

$\{y_i\}$, w, b, ξ

s.t. y_i ((w,x_i) +b>= 1-ξ_i ,ξ_i> 0

$\forall i{:}Y_I= 1,\sum_i \frac{yi+1}{2} >=1,$

$\forall i{:}Y_I= -1,\ y_i = -1,$

$y_i \in$ {-1,1} where w & b are weight vector and offset of SVM respectively, C is penalization parameter for misclassified instances &ξ_i are slack variable.

miSVM algorithm-

Initialize y_i =Y_I for i ∈ I:

repeat

compute SVM (w, b) for data with imputed labels;

compute outputs f_i = (w,x_i) +b for x_i in positive bags:

set f_i = sgn (f_i) for every i∈ I,Y_I =1;

foreach positive bag B1 do

if$\sum_{i\in I} \frac{1+yi}{2}$ = = 0 then

compute i^* = argmax i∈I f_i;

end

end

until impute labels do not change;

output: w, b

3. miSVM+PEDD classifier

The main drawback of miSVM is underestimation of positive instances in positive bags. This drawback is overcome by probability estimation method.

miSVM+PEDD algorithm-

Input: τ (probability threshold)Initialize $y_i = y_i' = Y_I$ for I ∈ I;

repeat

compute SVM (w, b) for data with imputed label

compute SVM class-conditional probability model (A,B);compute probability estimates P_i for x_i in positive bags;

foreach i ∈ I ,Y_I =1 do

if P_i>= τ then

set y_i =1;

else

set y_i= -1;

end

end

foreach positive bag B_I do

if$\sum_{i\in I} \frac{1+yi}{2}$ = = 0 then

compute i* = argmax i ∈ I Pi;

set yi* =1:
end
end
until impute labels do not change;
discard xi, yi ≠ yi* for every i∈I,Y_I =1;
output: w, b
The value of τ is taken as 0.5.

III. RESULT

TABLE 1 Image score of Zambia test set using SVM, miSVM and miSVM+PEDD methods

X-Ray No.	Image score		
	SVM classifier	**miSVM classifier**	**miSVM+PEDD classifier**
1	55	46	37
2	52	34	35
3	24	80	63
4	72	61	48
5	66	28	88
6	93	79	62
7	07	22	17
8	98	83	66
9	82	69	55
10	20	67	53
11	19	67	53
12	14	46	36
13	51	86	68
14	54	46	37
15	23	77	61
16	52	88	70
17	39	33	26
18	73	62	49
19	17	57	45
20	40	34	27
21	15	27	50
22	06	21	17
23	21	72	57
24	13	45	35

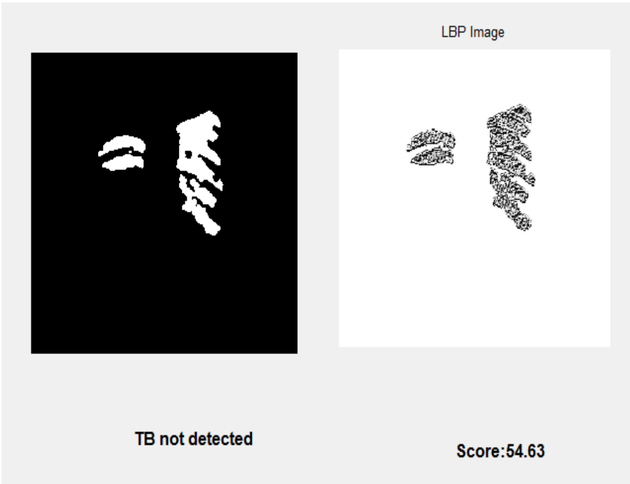

Fig.1 Image score of X1 using SVM method

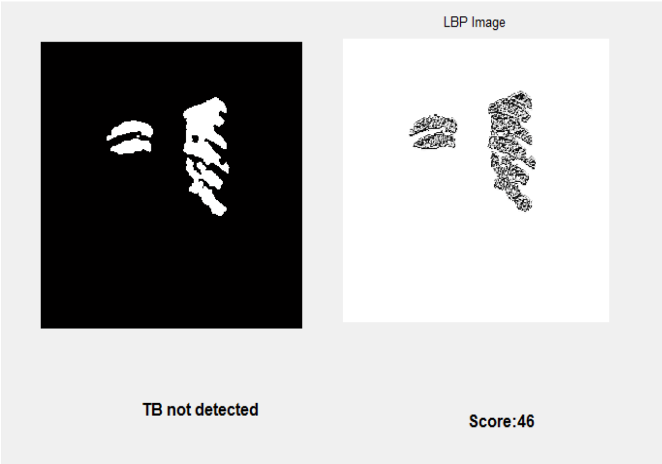

Fig.2 Image score of X1 using miSVM method

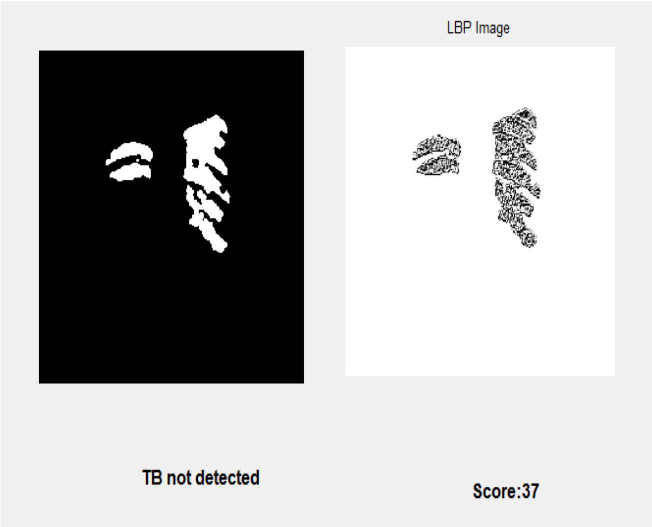

Fig.3 Image score of X1 using miSVM+PEDD method

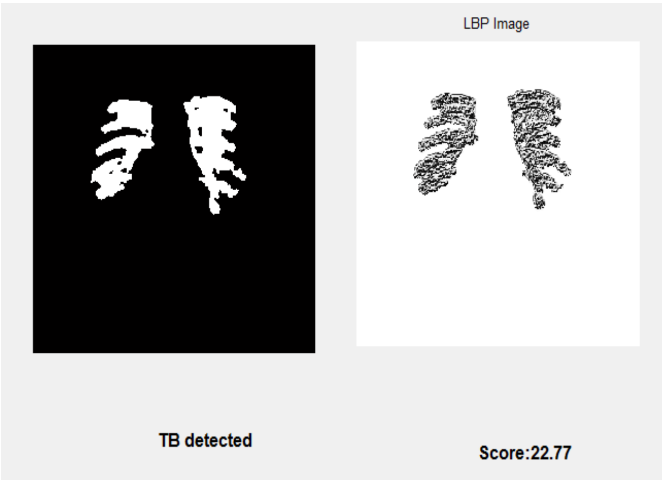

Fig.4 Image score of X15 using SVM method

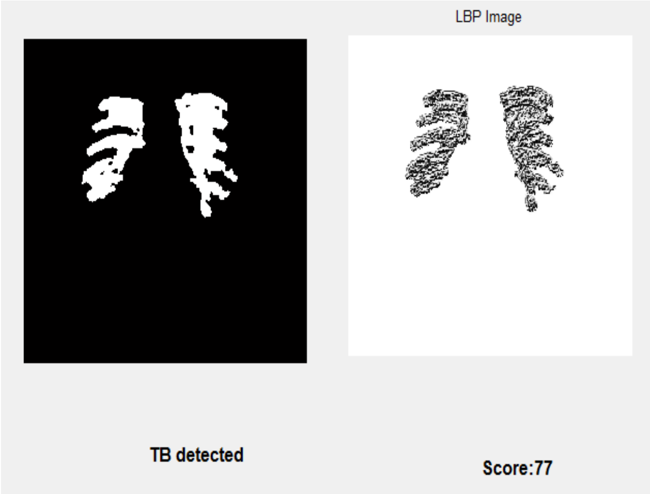

Fig.5 Image score of X15 using miSVM method

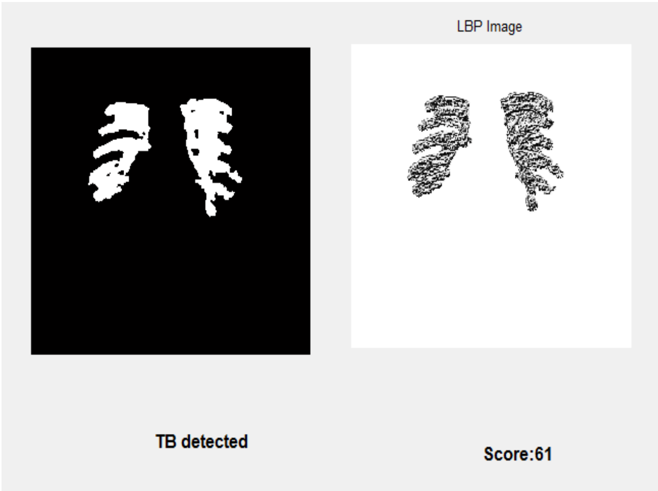

Fig.6 Image score of X15 using miSVM+PEDD method

Figure.1 gives image score 54.63 which we take as 55 and result 'TB not detected' using SVM method but for same image using miSVM and miSVM+PEDD methods (Figure 2 and Figure 3) show that 'TB detected'. Image scores of miSVM is varies due to use of set of patterns (bags). miSVM+PEDD insert probability which varies image score. miSVM and miSVM+PEDD shows image score 46 and 37. Figure 4 to Figure 6 shows all TB detected result with different image score. Though we say that image score below 50 gives 'TB not detected' and above it gives 'TB detected', SVM do not follow this rule. Image score may vary according to image. The result of test images is shown in above table 1.

TABLE 2CPU time required by SVM, miSVM and miSVM+PEDD to calculate image score

Database	Classification Method	CPU time(in seconds)
Zambia Test set	SVM	156
	miSVM	284
	miSVM+PEDD	448

CPU time using SVM method of Zambia test set is 156. miSVM gives more CPU time than SVM. miSVM+PEDD give highest CPU time 448 compared to other methods.

IV. CONCLUSION

In this paper, we have compared SVM, miSVM and miSVM+PEDD methods using image scoring method. This image score helps us for diagnosis of TB. Accuracy of SVM is 84%. miSVM gives accuracy 91% which is higher than SVM. miSVM+PEDD give accuracy 92% which is highest as compared to other methods.

V. REFERENCES

[1]. World Health Organization, Global Tuberculosis Report 2017

[2]. Jaime Melendez, Bram van Ginneken, Pragnya Maduskar, Rick H. H. M. et al. "A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays", *IEEE Trans. Med. Imag*.,vol.34,pp. 179-191, 2015

[3]. S. Jaeger, A. Karargyris, S. Candemir, et.al. "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, pp. 233–245, 2014

[4]. M.M. Dundar. G. Fung, B. Krishnapuram and R. B. Rao, "Multiple instance learning algorithms for computer-aided detection," *IEEE Trans. Biomed Eng.*,vol.55, pp. 1015-1021,2008

[5]. S.Andrews, I Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," *in Proc. Adv. Neural Inf. Process. Syst., 2003*, vol.15, pp. 561-568

[6]. Yixin Chen, James Z. Wang "Image categorization by learning and reasoning with region," *J.Mach. Learn.Res*., vol. 5, pp. 913–939, 2004

[7]. T.Ojala,M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEET. Pattern Anal.*, vol. 24, pp. 971–987, 2002

Mr.Onkar A. Badadare "Detection of Tuberculosis Using Novel Multiple Instance Learning Based Approach on Chest X-Rays., International Journal of Computational Engineering Research (IJCER), vol. 8, no. 7, 2018, pp. 54-59.