Automated Tuberculosis Detection and Analysis Using CXR's Images

Sana Fatima[1], Syed Irtiza Ali Shah[1*], Muhammad Zia Samad[2]
[1] National University of Sciences & Technology, Islamabad, Pakistan, Pakistani.
[2] National TB control program, Islamabad.

* Corresponding author. Email: irtiza_shah@gatech.edu
Manuscript submitted May 14, 2018; accepted June 5, 2018.
doi: 10.17706/ijcee.2018.10.4.284-290

Abstract: Infectious diseases are major health related problem in the world. Tuberculosis is included among the infectious, penetrating and prevalent disease caused by mycobacterium tuberculosis. Tuberculosis mostly effects lungs but also occurs in other parts of the body. The screening of tuberculosis is a tiresome task that requires highly trained workers. Screening and diagnosis of tuberculosis has been made possible by the discovery of X-ray radiology. The objective of this study is to summarize different automated tuberculosis detection approaches using chest X-ray and microscopic images. The paper present an automated system designed for screening of tuberculosis using CXR's. A total of 50 sample images were taken. Algorithm was implemented using MATLAB version 9.2. The interpretation of results was done using PASW statistics version 18. The accuracy of 92% was obtained for chest radiography algorithm that is above 90% so the algorithm can be used for screening of tuberculosis patients. Secondly, the digitize image is an important means for screening, diagnosis and treatment of tuberculosis. The proposed system will result in designing an efficient and accurate system with comparison to a manual system. In future these approaches can be used to other laboratory based techniques of tuberculosis screening e.g. smear microscopy and Drug Susceptibility test.

Key words: CAD, CXR, ROC, P-value.

1. Introduction

Image processing is one of the major field that is used in automating different techniques and procedures. Medical images processing is the field of image processing through which different medical diagnostic and treatment procedure are automated. Different medical procedures had been automatized previously and helped in the diagnosis and treatment of various diseases. The designed automated systems includes i.e. MRI, Computed tomography, CT scan, digital microscopy etc. Tuberculosis is an infectious disease that is transmitted through a medium that is a bacterium called mycobacterium tuberculosis. This research is based on automation for chest radiographs detecting tuberculosis. The CXR's samples are taken as a medium for screening of tuberculosis patients. The paper also discuss and analyze the techniques already designed for tuberculosis screening using CXR's and microscopy images.

2. Literature Review

This research study is a part of a review in which different automated techniques designed previously using chest radiographs and microscopy images samples are summarized [1].

3. Methods

3.1. Data Collection

The data of 50 patient's samples was obtained for designing chest radiography based detection system. Data was acquired from federal government services hospital Islamabad, Pakistan. Original results that were interpreted by medical professionals were taken as reference data. The digitize images of chest radiographs were taken for designing the algorithm.

3.2. Image Acquisition

The interpretation was done on the basis of width and height information however, the depth information was lost. The acquired images were having the width of 3000 pixels and length of 2400 pixels with the resolution of 72 dpi. The CXR's contains information for the detection of tuberculosis disease and other important lungs related diseases i.e. lung cancer, pneumonia etc.

3.3. Image Processing

Image processing was done using MATLAB R2017A with version 9.2 which provides the computing environment for implementation.

3.4. Watershed Segmentation and Thresholding

Chest radiograph is an important medium for screening of TB patients. Automation of chest radiography was done using watershed segmentation approach. Preprocessing of the images is done to focus the region of interest and to process images iteratively. In this step contrast of the images was also adjusted. Watershed segmentation and thresholding was done to separate the two lungs from the background. The lung objects are then detected and abnormality is identified. Post processing was done to categorize the images as major tuberculosis, minor tuberculosis from non-tuberculosis images.

3.5. Analysis and Interpretation of Results

The analysis of algorithm results was done using SPSS version 18.0 software [2]. Patients data was taken as variables for interpreting statistically that includes i.e. gender, age group, X-ray result, CXR Algorithm results, CXR algorithm outcome, Image types (Cavitation, Pleural effusion, Mediastinal hilar enlargement, milliary mottling, apical involvement, consolidations and air space infiltrates). Algorithm results were inserted in Microsoft excel 2013 containing the data of patients taken from hospital. PASW statistics version 18.0 was used for interpreting results. In addition chi-Square testing was performed for checking the association between the two techniques that is manual interpretation and algorithmic interpretation. Accuracy of the algorithm was also identified by ROC curve.

3.6. Computing Time

The time of processing for 50 sample images of CXR's was calculated by MATLAB component runtime 7.5.

4. Results

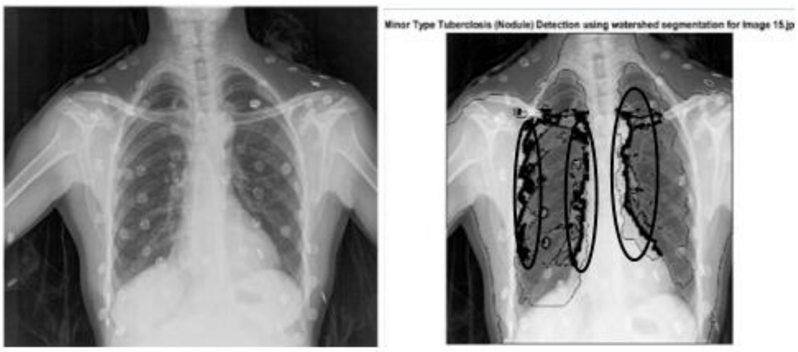
The CXR's (chest x-ray radiographs) contain useful information for the interpretation of different abnormalities. Tuberculosis is examined on CXR's by doctors and the process is automated by the proposed algorithm.

4.1. Automated Detection of Tuberculosis Using Watershed Segmentation Approach

The CXR images contain two dimensional representation of information about a three dimensional lung object for detecting tuberculosis. The important symptoms of tuberculosis identified in CXR's include pleural effusion, cavitation, consolidations, air space infiltrates, milliary mottling, apical involvement and mediastinal hilar enlargement. The symptoms are taken as data variables for comparison of CXR algorithm results with reference standard CXR data of the patients. Our algorithm is detecting tuberculosis in CXRs on the basis of symptoms present. The results differentiate major and minor TB cases from normal cases. The major tuberculosis is detected as the one in which three or more than three symptoms are present. The minor tuberculosis includes the one in which two symptoms are confirmed. However the images showing least chances of symptoms are shown as negative. In Fig. 1(a), a sample CXR image of a female is shown. From the reference standard data the results shows the presence of mediastinal hilar enlargement, apical involvement and pleural effusions. Some of the right lung region is also showing the symptoms of consolidations and milliary mottling. The algorithm detected this image as major type of tuberculosis which was correct according to the reference data. The output of the algorithm result is shown in Fig. 1(b) indicating abnormality. The regions in black indicate the major type of tuberculosis.

(a) (b)

Fig. 1. (a) This figure shows original chest radiograph image; b) This figure shows result for Fig. 1(a).



(a) (b)

Fig. 2. (a) This figure shows original chest radiograph image; (b) This figure shows result for Fig. 2(a).

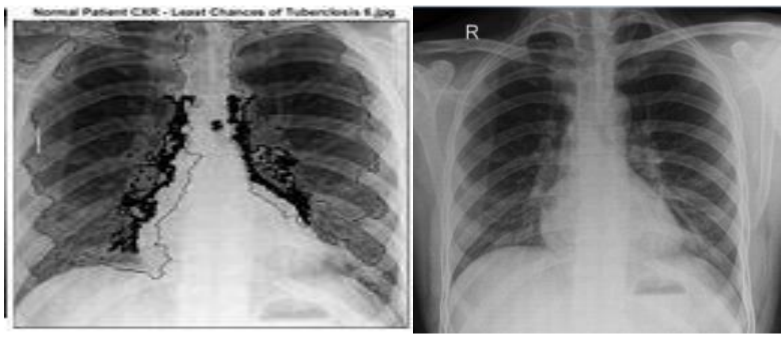


Fig. 3. (a) This figure shows original chest radiograph image; (b) This figure shows result for Fig. 3(a).

In Fig. 2(a), a CXR of a female is shown. The figure indicates that there are confirmed chances of apical region involvement and mediastinal hilar enlargement. The output is shown in Fig. 2(b) indicating the image as minor tuberculosis case. Our designed system has also detected normal images. The normal images are the one with least chances of the disease as shown in Fig. 3(a). This is a CXR of a male gender. There are some regions in Fig. 3(a), representing disease in the left lung but they are not confirmed. It is possible that these minor indications represented in the CXR are the tissues or some other sort of other disease e.g. pneumonia, heart related disease, tumor but are not confirmed indication for tuberculosis. The output from the algorithm is shown in Fig. 3(b), for normal image. The black color in figure shows that the CXR contained least chances for the presence of a disease.

The algorithm is detecting tuberculosis on the basis of pixels information of the lung object. Watershed segmentation iteratively scans the whole lung region and identifies those regions showing abnormality through a scanning window of fixed size. The threshold values are set which can help out in extracting the required information from the images.

5. Analysis of Data

The third part of this research study was to analyze the data using PASW statistics 18. The results of algorithms are compared with reference standard data. The number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are represented graphically with age group variable for Chest radiograph in Fig. 4(a). The constant bars are representing true positives and true negatives. However fluctuations in bars represents false positives and false negatives.

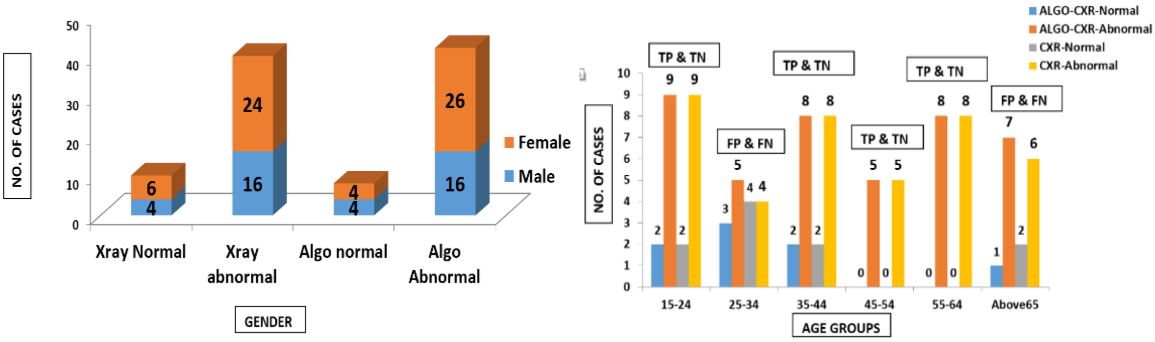


Fig. 4. (a) Graphical representation of reference standard data and algorithm data along with age group for chest radiography; (b) Graphical representation of reference standard data and algorithm data along with gender for chest radiography.

In addition to this, the variable gender is also taken with our results for Chest radiographs in Fig.4 (b). In the given dataset the number of females suffering from tuberculosis are greater as compared to male gender. The number of true positives, true negatives, false positives and false negatives are shown (see Table 1). This information is used to calculate sensitivity and specificity for the designed algorithm.

Table 1. Tabular Representation of Analysis and Pearson Chi-Square Test Results

Cross tabulation	X-rays Positive	X-rays Negative	Test Type	Remarks
Algorithm X-rays Positive	**39 (TP)**	**3 (FP)**	**Pearson Chi-Square test**	**As the p<05 which shows that there is no association between the two techniques of CXRs**
Algorithm X-rays Negative	**1 (FN)**	**7 (TN)**	**N**	**50**

The sensitivity of 98% and specificity of 70% was obtained in CXR tuberculosis detection algorithm. The positive likelihood ratio of 3.26 was obtained and the negative likelihood ratio of 0.03 was obtained for CXR tuberculosis detection system. The accuracy of 92% was obtained for chest radiograph images. The p value of less than 0.05 was obtained for the algorithm (see Table 1). This shows that there is no association between the proposed system and the standard reference system. This shows the independence of CXR's manual interpretation system from the proposed algorithm. The receiver operator curve (ROC) was generated that represents the accuracy of the designed algorithm. Area under the curve is the region for finding the accuracy of any system. The area under the curve of 0.838 was obtained for chest radiography tuberculosis detection algorithm as shown in Fig. 5.

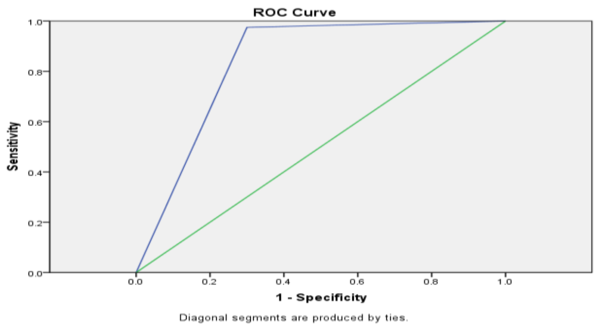


Fig. 5. This figure shows ROC interpretation for chest radiography algorithm. (Null hypothesis shown by green and alternative hypothesis by blue).

The sensitivity of 0.975 and specificity of 0.3 was taken as cut off point for detecting the abnormality correctly. The value above or below this cutoff value will detect either sensitivity or specificity correctly. The cut off value is the point for obtaining the desired sensitivity or specificity of the algorithm. The execution time for the algorithm was calculated using MATLAB command (Run and Time) for checking the speed. The profile time indicates the total time for execution of algorithm. Execution time for CXR abnormality detection algorithm was 564 Seconds for 50 CXRs sample images.

6. Discussion

The paper is a continuation of a review already done for already designed automated screening techniques for tuberculosis [1]. Different approaches designed have some advantages and limitations. From these techniques an important system is proposed for automated screening using CXR's samples. Watershed segmentation approach was followed for chest radiography algorithm from the work done earlier and lungs were isolated for identifying tuberculosis [3]. However in the previous research no such accuracy was mentioned. Previous work was done for other abnormalities of the lungs regions too. Our proposed system would focus on tuberculosis and had shown reliability of use with accuracy of 92% which is greater than other algorithms designed [1]. Another automated system was designed named TBDx in which the sensitivity of 62% and specificity of 100% was obtained [4]. However for chest radiographs sensitivity is more important as compared to specificity as we are not counted AFB's in chest radiographs but detecting the abnormalities. The designed system is best for identifying abnormality because of good sensitivity. A system was designed with automated tuberculosis detection procedure which was used for smear microscopy but it incorporated manual segmentation [5].Different variables from reference standard data were compared with algorithm results. Interpretation of algorithm was compared with reference standard technique results. The numbers of true positive and true negative are greater than false positive and false negative so our designed algorithm is accurate in detecting tuberculosis. The sensitivity and specificity is calculated for checking the usefulness and correct testing by the algorithm. Sensitivity of algorithm was greater as compared to specificity which is good in one way and a limitation on the other hand. CXR's detection algorithm will identify minor detail from the chest radiograph but there are chances that our algorithm will detect diseases other than tuberculosis in the lungs at the time of execution. The positive likelihood of greater than one indicates correct results. Similarly, the negative likelihood ratio of less than one is obtained which also shows correctness of the proposed system. The accuracy of greater than 90% shows that we can use the proposed system for detecting tuberculosis effectively. Pearson Chi-Square test interpretations shows that there is no association between the designed algorithm and standard reference technique (manual system) available. The P-value of less than 0.05 was obtained which shows that the techniques are independent of each other and the null hypothesis was rejected. According to null hypothesis the techniques are dependent on one another. However, the alternative hypothesis suggests that they are independent of each other. The alternative hypothesis was accepted for chest radiography algorithm. The receiver operator curve was interpreted for the algorithms for analyzing the accuracy. According to null hypothesis no sensitivity or specificity will be obtained where Ho will be equal to 0.5. The null hypothesis was rejected as shown by blue lines in Fig. 5 and the area under the curve of 0.838 was obtained for chest radiography algorithm. The cut off value was calculated for obtaining the accuracy at which both sensitivity and specificity will be good. The cutoff points are discussed earlier in results. Computing time for the proposed chest radiograph algorithm was greater because it contains large amount of pixel information for interpretation. Radiologists would be able use the proposed algorithm in the hospital for identifying the disease. Reports will be generated and stored in folder iteratively. They will be used by doctors and medical specialists in diagnosis and treatment of the disease.

7. Conclusion

The presence of abnormality in the proposed Chest radiography algorithm possessed higher sensitivity and therefore would assist in differentiating between minor and major tuberculosis. The accuracy of the proposed algorithm was good so it can be implemented for screening. Computation time for chest radiography algorithm is more however, it could reduce human fatigue by processing all the images available iteratively automatically. However, the confirmation for the presence of the disease is not much clear from one type of test. Therefore, different types of tests are performed for reliable TB identification.

8. Recommendations & Future Work

Chest radiograph detection system should be made more specific so other features of lungs should be also be taken into consideration which can detect abnormalities other than tuberculosis e.g. cancer, heart diseases and pneumonia. Secondly, the system can be designed for other techniques of tuberculosis screening too that include smear microscopy and drug susceptibility test.

Acknowledgement

We are thankful to all of those with whom we have had the pleasure to work during this project. We are especially appreciative to the staff of Polyclinic Hospital, Islamabad for helping me out in data collection procedure.

References

[1] Sana, F. & Shah, S. A. I. (2017). A review of automated screening for tuberculosis of chest Xray and microscopy images. *International Journal of Scientific & Engineering Research, 8(10)*, 2229-5518.

[2] How to Use SPSS: Calculating Sensitivity, Specificity & Likelihood Ratios. (2017, December).

https://www.youtube.com/watch?v=H0olYl2mJO0

[3] Le, K. (2011). A design of a computer-aided diagnostic tool for chest x-ray analysis. *International Journal of Computer Science & Information Technology, 3*, 212-222.

[4] Shah, M., Mishra, S., Yadav, V. K., Chauhan, A., Sarkar, M., Sharma, S. K., & Rout, C. (2017). Ziehl-Neelsen sputum smear microscopy image database: A resource to facilitate automated bacilli detection for tuberculosis diagnosis. *J Med Imaging (Bellingham), 4(2)*, 027503.

[5] Nabeta, P., Havumaki, J., Ha, D. T. M., Caceres, T., Hang, P. T., Collantes, J., *et al.* (2017). Feasibility of the TBDx automated digital microscopy system for the diagnosis of pulmonary tuberculosis. *PLoS ONE 12(3)*, e0173092

Sana Fatima was born in Pakistan. Sana earned her BS degree from International Islamic University, Islamabad in the field of bioinformatics in 2013. Sana did her MS from National University of Sciences & Technology, Islamabad in the field of biomedical sciences in 2017. She has been working as a screener in National TB Control Program since 2017 and responsible for screening tuberculosis cases from the presumptive cases. She has been working as a data entry officer in National TB Control Program since 2014-2017. As a data entry officer, she was responsible for collecting and analyzing data of tuberculosis patients.

Ms. Fatima published a review article on “A review of automated detection techniques and their comparative analysis” in International Journal of Scientific and Engineering Research.

Ms. Fatima was acknowledged in the paper titled “Prevalence of TB-HIV Co-Infection in Pakistan” that was on tuberculosis research.

Syed Irtiza Ali Shah is a professor in National University of Sciences and Technology (NUST) Pakistan. He did his PhD from Georgia Institute of Technology USA in 2010, prior to which he did the MS in mechanical engineering (robotics) in 2008 and the MS in aerospace engineering (automatic controls) in 2007 from the same institute. He also did the MS in fluid mechanics from NUST in 2004 and the MSc in computer science from CoCS Pakistan in 2000. He has more than 75 publications to his credit and has completed about 50 funded projects related to vision systems development for medical, robotic and aerospace applications.

Muhammad Zia Samad was born in Pakistan. Zia earned his master degree from the Islamia University Bahawalpur in the field of statistics in 2001. Currently Muhammad Zia is doing the MPhil in statistics from Allama Iqbal Open University, Islamabad, Pakistan.

He has been working as a surveillance officer and data analyst in National TB program since 2013 and responsible for developing, managing, supervising and produce the statistical analyses of the districts and provincial TB data. From 2011 to 2013, He has been the data management coordinator in Pakistan TB Disease Prevalence Survey. Mr. Samad published an article entitled “Alarming Rates of Attrition among Tuberculosis Patients in Public-Private Facilities in Pakistan”.