International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 11 | Nov 2020 www.irjet.net p-ISSN: 2395-0072

Automated Detection of Pulmonary Tuberculosis in Digital Chest Radiographs Using Image Processing Techniques and a Statistical Approach

A. J. S. P. Nileema[1], S. Kulatunga[2], S. H. Palihawadana[3]

[1]PG Student, Postgraduate Institute of Sciences, University of Peradeniya, Peradeniya, Sri Lanka
[2]Senior Lecturer, Department of Radiography/Radiotherapy, Faculty of Allied Health Sciences, University of Peradeniya, Peradeniya, Sri Lanka
[3]Consultant Radiologist, Department of Radiology, National Hospital for Respiratory Diseases, Welisara, Sri Lanka

---

Abstract - *The objective of the study was to develop a computer aided detection (CAD) system for the classification of digital chest X-rays (CXRs) into pulmonary tuberculosis (PTB) - negative (PTBN) and PTB - positive (PTBP) classes with the aid of image processing techniques in MATLAB and a statistical approach. 200 digital CXRs were compiled from National Hospital for Respiratory Diseases, Welisara, Sri Lanka. CXRs were pre - processed to remove identification details. Lung fields were segmented with a thresholding approach. Segmented images were divided into four quadrant images. Contrast, correlation, homogeneity, energy, entropy and maximum probability texture features were extracted from quadrant images using gray level co-occurrence matrix method. Feature values were analyzed for descriptive statistics and normal distribution using SPSS. Images were classified manually into PTBP and PTBN classes depending on radiologists' interpretations. An independent sample T- test was conducted to compare the feature values between the two classes and a classification algorithm was defined using features with statistical significance to classify the quadrant images of all CXRs into either PTBN or PTBP. The CAD system showed that contrast, correlation, energy, entropy and maximum probability can be used in CXR classification for PTB diagnosis as these features have shown statistically significant difference between PTBP and PTBN CXRs. The developed CAD system for classification of CXRs has shown a high sensitivity of 92% which makes it an efficient CAD system and with a modest specificity of 70%.*

Key Words: Image Processing, Computer Aided Detection, Pulmonary Tuberculosis, Chest Radiographs, MATLAB, Classification, Lung Segmentation, GLCM

1. INTRODUCTION

Despite the significant progress in diagnosis and effective treatment methods to eradicate the disease, tuberculosis (TB) remains one of the world's deadliest communicable diseases according to the WHO. TB is one of the top 10 causes of death and the leading cause from a single infectious agent [1]. Especially in developing countries TB is known as the leading cause of morbidity and mortality. Yet TB has received less attention from general public and media in spite of these findings. Globally millions of people continue to fall sick with TB each year [1].

TB is caused by the bacillus *Mycobacterium tuberculosis* (*MT*). It is an airborne disease and recognized as a highly infectious disease due to its nature of transmission. Lungs are the main sites of infection for *MT* bacillus due to its preference for high Oxygen environment and when the lungs are infected with *MT* bacillus it develops pulmonary tuberculosis (PTB) [2]. When it affects other areas of human body the disease is known as extra pulmonary TB. Patient with active PTB may be symptomatic, have mild or progressive dry cough, or present with multiple symptoms including fever, fatigue, night sweats, loss of appetite, weight loss, cough with blood contained sputum and significant nail clubbing [1, 3].

PTB is diagnosed using a full combination of medical history, physical examination, laboratory tests and imaging studies [4]. A chest X-ray (CXR) and multiple sputum cultures are typically included in the initial assessment for PTB. Sputum smears microscopy remains still as the most widely available method for diagnosing PTB as it is really inexpensive [5]. However, culturing of *MT* bacillus has been established as the gold standard for diagnosis of PTB since it was developed [6, 7]. Use of conventional CXRs has been identified as an essential tool in diagnosing PTB at earlier stages since the beginning of the twentieth century in low and high burden settings [4, 8, 9].

A prompt and accurate diagnosis followed by provision of treatment in line with international standards prevents deaths and the onward transmission of the infection can be curtailed in improving the community public health. Automated DNA based analysis tests are quick and gives accurate results. However, those tests are expensive making it unsuitable for large population screening [10]. With the introduction of digital radiography (DR) the use of CXRs in diagnosing PTB has further extended leveraging its advance features. CXRs meet the requirement for a screening or triage process due to its high sensitivity for PTB and it is a quick and a reliable diagnosis method. CXRs can serve both as a screening and a measurement tool to identify pulmonary diseases including PTB [11]. Thus CXRs are

considered as an essential inexpensive component of the early diagnosis of PTB [11, 12].

Further, existing literature have validated that CXR is the most sensitive screening tool for PTB diagnosis [12-15]. Despite these advantages, CXR interpretation is still subjected to human errors and depends on the expertise of the radiologist as well as results in high over and under diagnosis. Further, inter - reader variation is also present when manual reading of CXR is employed. Since 1963 [16, 17], numerous studies have actively researched for fully automated PTB diagnostic systems by CXRs, yet there is no standard [18-26]. Nevertheless, a computer aided detection (CAD) technology could help reduce inter - reader variability and delays in reporting CXRs when skilled radiologists are limited especially in high TB prevalence countries with fewer resources.

Thus developing a CAD system for diagnosing PTB in digital CXRs would be of a great benefit to health care system. In addition availability of a CAD system would allow the clinicians in a mobile screening program to intentionally over - read the CXRs of whose findings are suspicious for PTB or whose CXRs are minimally abnormal. Further, WHO *global TB Report 2018* stated that still there is the requirement for new intensified research studies towards developing a standard for detecting PTB using digital CXRs [3]. Therefore in this research study the objective was to develop an automated system for detecting PTB using digital CXRs with the use of digital image processing techniques and a statistical approach.

2. MATERIALS AND METHODOLOGY

The study was conducted following the ethical approval from Ethical Review Committee of Faculty of Faculty of Allied Health Sciences, University of Peradeniya. Permission was also obtained from the Director; National Hospital for Respiratory Diseases (NHRD), Welisara, Sri Lanka.

A cross sectional descriptive study was carried out using CXRs collected from outpatient department, NHRD. All CXRs taken between 01st of September to 30th September 2019 were included in the study sample whereas CXRs which were of suboptimal diagnostic quality caused by overexposure, underexposure, region of interest cut off, rotation of the patient, malpositioning, improper breathing and presence of artifacts were excluded from the study. A convenient sample of 200 CXRs was compiled as the study sample of images. All CXRs of study sample were captured with the DR system at outpatient department, NHRD which comprised of Shimadzu static X-ray machine in cooperated with a Canon CXDI- 401C COMPACT digital image detector in a 12-bit gray scale.

The system for automatic detection of PTB using digital CXRs was designed as a set of cascaded components. These steps consisted of pre-processing of CXRs, lung region segmentation, dividing the lung fields into quadrant images (QI), feature extraction, statistical analysis for classification decision and classification of QIs. All the steps are described in this chapter and an overview of the steps is presented in Figure 1. The collected data set were first studied and all the CXRs which had any of the exclusion criteria were excluded from the sample. Then the compiled 200 CXRs were anonymized to remove patient identification details in CXRs using the Radiant DICOM viewer before using in the study.

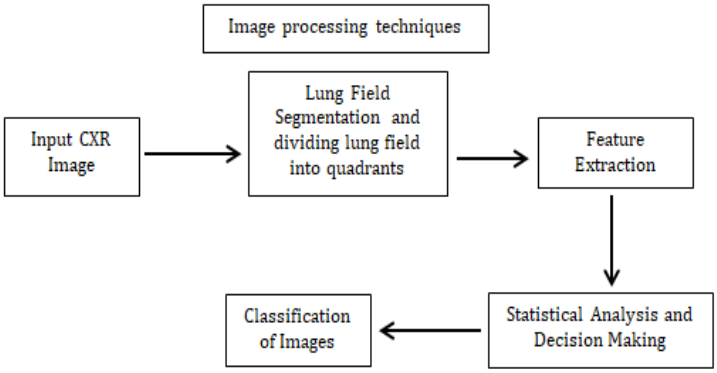

Fig - 1. An overview of the CAD system.

2.1. Lung Region Segmentation and Dividing into Quadrants

A vital first step in any CAD program using CXRs is the segmentation of lung fields from its background. Thus in this step both right and left lung volumes were extracted using the image processing algorithms in MATLAB. As an initial step the preprocessed CXRs were named with a specific pattern for the data set and stored in a separate folder. Then the series of CXRs were imported to the MATLAB workspace by assigning it as a variable. The first step was to convert the true color CXR into a gray scale image (GSI). It was then converted to a binary image (BI) format by thresholding. Subsequently unwanted light structures connected to border of BI were suppressed using *imclearborder* function. After that *bwareaopen* was used with an assigned pixel value to remove all objects less than the assigned pixel value in the binary image. In the resulted BI any of the unoccupied area which was holes was filled using a filling algorithm and this was used as the binary mask image (BMI). Finally, the original GSI was masked with the BMI resulted from previous step and it produced a gray scale segmented lung image (SLI). Lung segmentation process using a masking approach employed is demonstrated in Figure 2.

The SLI was then divided into 4 QIs; right upper quadrant (RUQ), left upper quadrant (LUQ), right lower quadrant (RLQ) and left lower quadrant (LLQ). This was achieved by defining the dividing points for each QI in the developed MATLAB algorithm; to divide the segmented right and left lung fields in such a way that there will be equal halves for each side of the lung creating the four quadrants.

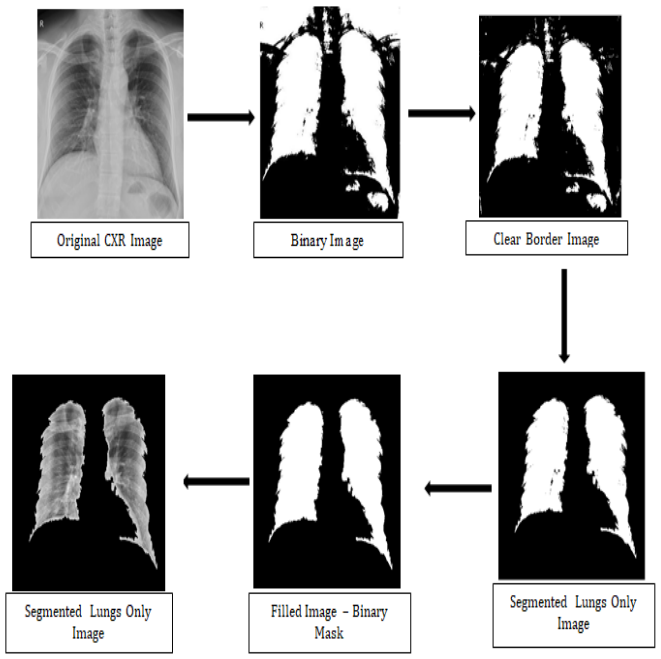

Fig -2. Illustration of the lung segmentation algorithm.

2.2. Feature Extraction

To find out the PTB-negative (PTBN) and PTB-positive (PTBP) variances in the segmented lung fields six texture features (TFs) including contrast, correlation, energy, homogeneity, entropy and maximum probability were extracted from each QIs using GLCM method in MATLAB. All the feature values were calculated for each QI separately and were automatically recorded in Microsoft Excel data sheets accordingly.

2.3. Statistical Analysis and Classification Decision

Descriptive statistics were calculated for all the extracted variables. Each feature value was tested for standard normal distribution using skewness and kurtosis values as the statistical independent sample T-test requires normally distributed data sets of variables. CXRs were manually classified into PTBN and PTBP categories according to the radiologists' interpretations and TF data were recorded separately for each category of QI respectively. Then an independent sample T - test was conducted for the normally distributed TFs at 95% confident interval for QIs to find any existing statistical significant difference (SSD) in features between PTBP and PTBN classes of images. Results were recorded accordingly. Such features resulted with a SSD in mean values between PTBN and PTBP images were identified as the parameters of image classification. The range of minimum and maximum value was specified for these identified parameters for future use in deciding if the testing image falls in the PTBP category.

2.4. Classification of CXRs

An algorithm for classification was defined using the range of SSDF values obtained to identify and classify the QIs into PTBP - "Abnormal" and PTBN - "Normal" classes. The resulted ranges of SSDF values when a QI is showing PTBP findings were used as the criterion to define the "Abnormal" category. The algorithm was defined in such a way that; if the feature values are within the defined range the algorithm will label the respective QI as "Abnormal" in red color and the border of the QI will be outlined in red color. When the TFs are out of the defined range for the QI being tested it will be labeled as "Normal" in blue color and there will not be any changes to the existing border of the QI.

3. RESULTS

All the 200 CXRs of the data set were successfully processed to segment the lung fields distinctly out of the entire CXR with the developed algorithm. In the subsequent step the segmented lung images were divided into QIs for further in-depth analysis of PTBP and PTBN variations throughout the lungs in the four main lung regions using the extracted TF values. This dividing into QI was done to achieve an in detailed analysis of the CXRs such as which part of the lungs are mostly being affected by the disease and is the distribution of the disease is equal in the entire lung field.

According to the descriptive statistics five TFs including contrast, correlation, energy, entropy and maximum probability showed a normal distribution. In manually dividing the CXRs into PTBN and PTBP classes 103 CXRs reported as "Abnormal" and 97 CXRs reported as "Normal" were identified. All the feature data were tested and interpreted at 95% of confident interval, and p values equal or less than 0.05 ($p < 0.05$) is interpreted to have SSD in values between the two classes tested.

When comparing with PTBN CXRs, contrast showed a lower statistically significant mean values (SSMVs) in all four QIs of PTBP class with $p < 0.001$ (82 ± 64, 156 ± 119 (RUQ), 64 ± 14, 226 ± 131 (LUQ), 87 ± 70, 182 ± 128 (RLQ) and 100 ± 79, 203 ± 157 (LUQ)). Correlation showed higher SSMVs in PTBP CXRs of all four QIs with $p < 0.001$ (0.98 ± 0.01, 0.97 ± 0.02 (RUQ), 0.98 ± 0.01, 0.97 ± 0.02 (LUQ), 0.99 ± 0.01, 0.97 ± 0.02 (RLQ) and 0.98 ± 0.01, 0.96 ± 0.02 (LLQ)).

Energy had higher SSMVs in RUQ; $p < 0.001$ (0.41 ± 0.10, 0.36 ± 0.08) and in LUQ; $p = 0.001$ (0.43 ± 0.09, 0.39 ± 0.07) in PTBP class. Entropy resulted with lower SSMVs in PTBP CXRs in RUQ (3.35 ± 0.44, 3.59 ± 0.41) and LUQ (3.22 ± 0.43, 3.40 ± 0.47) with $p < 0.000$ and $p = 0.004$ respectively. Maximum probability had higher SSMV only in RUQ with $p < 0.001$ (0.63 ± 0.06, 0.59 ± 0.05) PTBP CXRs. Hence it can be said that both right and left lung volumes are affected by *MT* bacillus and shows PTBP findings according to the results obtained from the study. Also spread of the infection on both right and left sides is asymmetrical. Further, the effect of *MT*

bacillus causing PTB affects greater disparities in all feature values tested between right and left lung volumes.

In the final step the value ranges shown in the Table 1, were used to define the classification algorithm. All 200 CXRs were processed successfully for classification process. Figure 3., shows examples for the classified QIs and their respective input CXRs, segmented lung image and the QI.

The developed CXR classification algorithm showed a sensitivity of 92% and specificity of 70%. Further the developed classification algorithm resulted with a positive predictive value (PPV) of 76% and a negative predictive value (NPV) of 89%.

Table -1. Value ranges of TFs used for the classification of QIs.

Quadrant	Feature	Minimum Value	Maximum Value
RUQ	Contrast	18	146
	Correlation	0.97	0.99
	Energy	0.31	0.51
	Entropy	2.91	3.79
	Maximum Probability	0.57	0.69
LUQ	Contrast	50	78
	Correlation	0.97	0.99
	Energy	0.34	0.52
	Entropy	2.79	3.65
RLQ	Contrast	17	157
	Correlation	0.98	1.00
LLQ	Contrast	21	179
	Correlation	0.97	0.99

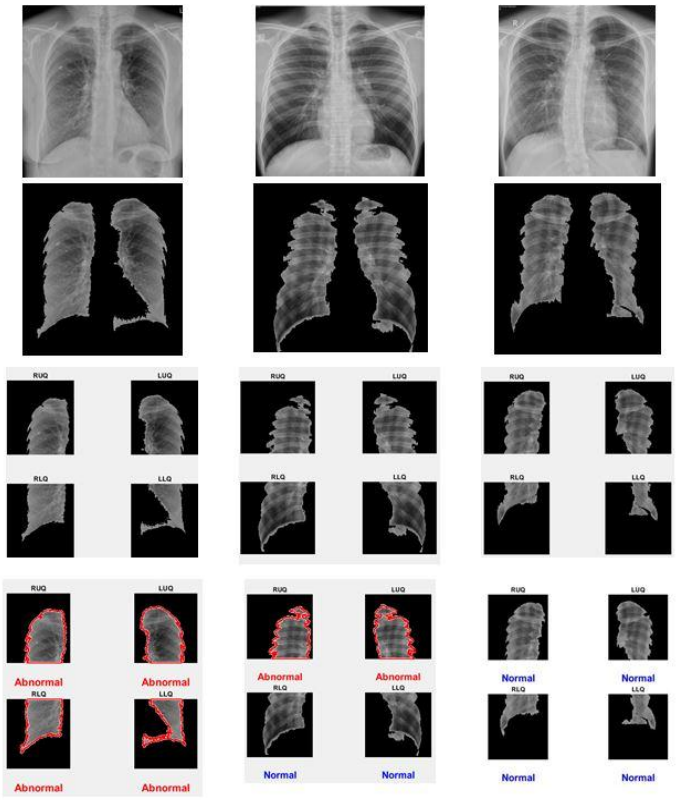

Fig -3: Illustration showing the resultant images from the data set with the final classification results with first row from top: original CXRs used in the study, second row from top: segmented lung field images, third row from top: quadrant images and fourth row from top: shows the classified quadrant images.

4. CONCLUSIONS

Lung segmentation was achieved for most of the CXRs successfully to encompass the best possible lung field. Therefore this algorithm can be concluded as a beneficial technique in lung segmentation which can be used for further studies with further modifications. GLCM method was successful in feature extraction of digital CXRs. It can be concluded that contrast, correlation, energy, entropy and maximum probability features can be used in CXRs classification for PTB diagnosis since there were a statistically significance difference between PTBP and PTBN CXRs in these features. The developed classification algorithm had resulted with a high sensitivity of 92% which makes it an efficient CAD system and with a modest specificity of 70%. Finally it can be asserted that both right and left lung volumes are affected by *MT* bacillus and shows PTBP findings. Also spread of the infection on both right and left sides is asymmetrical. Further PTB affects variations in most of the feature values tested between right and left lung volumes.

REFERENCES

[1] Global tuberculosis report 2019. Geneva: World Health Organization; 2019.

[2] Cole E.C., Cook C.E., "Characterization of infectious aerosols in health care facilities: an aid to effective engineering controls and preventive strategies". American Journal of Infection Control., 1998; **26** (4): 453–64.

[3] Global tuberculosis report 2018. Geneva: World Health Organization; 2018.

[4] Sameer A., Sema C., Stefan J., Alexandros K., Santosh K.C., Szilárd V., Zhiyun X., Automated Detection of Lung Diseases in Chest X-Rays, A Report to the Board of Scientific Counselors April 2015, An intramural Research division of the U.S. National Library of Medicine.

[5] Boehme C.C., Saacks S., O'Brien R.J., The changing landscape of diagnostic services tuberculosis, Semin Respir Crit Care Med, 2013 Feb;34(1):17-31.

[6] Perkins M.D, Roscigno G., Zumla A., Progress towards improved tuberculosis diagnostics for developing countries, Lancet Lond Engl,2006; 18;367(9514):942-3.

[7] Singhal R., Myneedu V.P., Microscopy as a diagnostic tool in pulmonary tuberculosis, Int J Mycobacteriology, 2015; 4(1):1-6.

[8] Lerner B.H., The perils of "X-ray vision": How radiographic images have historically influenced perception, Perspect Biol Med, 1992; 35(3):382-97.

[9] Bakari M., Arbeit R.D., Mtei L., Lyimo J., Waddell R., Matee M., Cole B.F., Tvaroha S., Horsburgh C.R., Soini H., Pallangyo K., von Reyn C.F., Basis for treatment of tuberculosis among HIV-infected patients in Tanzania:the role of chest x-ray and sputum culture, BMC Infect Dis, 2008; 6;8:32.

[10] Global tuberculosis report 2014. Geneva: World Health Organization; 2014.

[11] World Health Organization. Tuberculosis Prevalence Surveys: A Handbook. 2011.

[12] World Health Organization, Chest Radiography in Tuberculosis Detection, Geneva, 2012, (WHO/HTM/TB/2016.20.)

[13] Maduskar P., Muyoyeta M., Ayles H., Hogeweg , Peters-Bax L. , van Ginneken B., Detection of tuberculosis using digital chest radiography: automated reading vs. interpretation by clinical officers, int j tuberc lung dis, 2013; 17(12):1613–1620.

[14] Satheeshkumar K.G., Alex Noel J.R., Developments In Computer Aided Diagnosis Used For Tuberculosis Detection Using Chest Radiography: A Survey, ARPN Journal of Engineering and Applied Sciences, 2016; 11:5530-5539.

[15] Stefan J., Alexandros K., Sema C., Jenifer S., Les F., Sameer A., George T., Automatic screening for tuberculosis in chest radiographs: a survey, Quant Imaging Med Surg , 2013;3(2):89-99.

[16] Lodwick, G.S., Computer-aided diagnosis in radiology: A research plan. Invest. Radiol. 1:72–80, 1966.

[17] Lodwick G.S., Keats T.E., Dorst, J.P., The coding of Roentgen images for computer analysis as applied to lung cancer. Radiology,185–200, 1963.

[18] Stefan J., Alexandros K., Sameer A., George T., Detecting Tuberculosis in Radiographs Using Combined Lung Masks, in Proc of the Annual International Conference of the IEEE EMBS'5, 2012; 4978-4981.

[19] Jen H.T., Rajendra Acharya U., Collin T., Thomas A. K., Choo M.L., Computer-Assisted Diagnosis of Tuberculosis: A First Order Statistical Approach to Chest Radiograph, J Med Syst,2012; 36:2751–2759.

[20] Hogeweg L., Mol C., de Jong P.A., Ayles H., van Ginneken B., Fusion of Local and global detection systems to detect Tuberculosis in chest radiographs, Med Image Comput Comput Assist Interv, 2010; 13 (Pt 3): 650-657.

[21] Patil S.A., Texture Analysis Of Tb X-Ray Images Using Image Processing Techniques, Journal of Biomedical and Bioengineering, 2012;3(1):53-56.

[22] Chetan C. P., Ganorkar S. R., Tuberculosis Screening Using Digital Image Processing Techniques, International Research Journal of Engineering and Technology (IRJET), 2016;3(7):623-627.

[23] Tao Xu. (2013), Intelligent CAD System for Infectious TB Detection on Chest Radiographs, Ph.D. Thesis, University of Alberta, Canada.

[24] Hrudya D., Ajay N., An Efficient Detection of Tuberculosis from Chest X-rays, International Journal of Advance Research in Computer Science and Management Studies, 2015; 3(5): 149-154.

[25] Beaulah J. R., Balasubramanian R., Anbarasa P., An Efficient Feature Extraction Method for Tuberculosis detection using Chest Radiographs, International Journal of Applied Environmental Sciences, 2017;12(2):227-240.

[26] Kamble P. A., Anagire V. V., Chamtagoudar S. N., CXR Tuberculosis Detection Using MATLAB Image Processing, International Research Journal of Engineering and Technology (IRJET), 2016;3(6):2342-2344.