Computer Aided Tuberculosis Detection, A Review

Pragya Shukla[1], Jasleen Saini[2], Dr. B. S. Saini[1]

[1] Dept. of Electronics and Communication, Dr. B.R. Ambedkar National Institute of Technology Jalandhar

[2] Dept. Of computer science, UIET, Chandigarh, Punjab University

*Corresponding author

doi: https://doi.org/10.21467/proceedings.114.20

Abstract

This paper aims at presenting a complete picture of advances till now in the field of computer-aided detection of Pulmonary Tuberculosis using Chest X-ray Images. Advances are analyzed in chronological order as they happen and are divided into three phases in which technology shifted into new paradigms. Study concludes that although techniques that use Machine learning based methods for segmentation and classification are prevailing for the moment in terms of flexibility for very particular feature extraction in borderline cases where probabilistic methods can be tweaked according to requirements and accuracy, Deep Convolutional Neural Network based technique will secure higher standings as the computational capability and dataset management improves. Finally, briefly an attempt at using visualization techniques for borderline cases is discussed.

1 Introduction

Tuberculosis is a highly infectious bacterial disease, caused by a bacteria called *Mycobacterium Tuberculosis.* Tuberculosis bacteria can affect any organ in the host, but it is most common in lungs, which is called Pulmonary Tuberculosis. According to WHO report 2020[1], 10 million people got reported with Tuberculosis infection in 2019, and a total of 1.4 million people died in 2019 of tuberculosis infection, out of which 208000 were also HIV positive together with Tuberculosis infection. This huge number makes Tuberculosis as one of the deadliest diseases in the world, reported under top ten major causes for death in WHO report. Agonizing figures of data not only points out the prevalence of the problem itself but also the urgent need for action in terms of early detection, treatment and rehabilitation. Tuberculosis is a majorly treatable disease today and early detection plays an important role in treating disease completely without losing life expectancy years. X-ray Imaging is a prevalent method used for Lung TB detection, X-ray Imaging can be used for diagnosis as well as for tracking progress of treatment. Although Chest X-ray has been an important tool for diagnosis of pulmonary Tuberculosis but the scarcity of expert radiologists in the area pushed the need for computer aided techniques to not only diagnose the disease but do it at a low cost.

Research so far can be roughly divided into three phases in which different approaches towards computer aided diagnosis developed with the advances in Imaging Technologies and computing capabilities. In the first phase in early 2000s, methods of segmentation of an image were developed, they took pixel by pixel approach for local texture based analysis [2].

In the second phase, advances in machine learning prevailed with automatic segmentation, knowledge guided models[3], graph-cut segmentation methods[4] were developed, usage of algorithms such as gradient histogram, decision tree, GLCM, k-neighbours for feature extraction came into focus, and popular machine learning based tools such as SVM and PCA algorithms were used for classification. Second phase was highly influenced by advances in machine learning and all the major steps in Computer aided Tuberculosis diagnosis namely segmentation, feature extraction and classification changed. Accuracy of Computer Aided

Proceedings DOI: 10.21467/proceedings.114; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-947843-8-8

diagnosis improved and became on par with humans. For the first time in this phase, Computer aided diagnosis became a viable option and WHO encouraged the idea. Few start-ups, such as qure.ai in India, came into existence as the research progressed significantly enough to promise an alternative to human radiologist.

Finally, the third phase of Deep learning convolutional neural networks emerged, it intersected with the machine learning era for a while before dominating the research arena completely today. Today research is dominated by unique architectures of Convolutional Neural Networks. Tweaks in filters used for convolutional layer, pooling layer and fully connected layer are being tested for better accuracy by many researchers such as [18]. Since CNN are computationally expensive, attempts at transferring learning from popular already existing trained models of non-biomedical image datasets is also a topic of research today. Deep learning based CNN models are already on par with Machine learning based models as far accuracy, sensitivity and specificity parameters are concerned and CNN models are ought to outperform machine learning models in near future because of the very nature of CNN models as the dataset quantity and quality along with computational capabilities increases.

Phase. 1

With the advances in digital Image processing, approaches towards computer aided detection started in this phase [20],[21], [2], [19]. Although the idea of computer-aided started taking its root but accuracy for methods wasn't on par with humans yet. Methods of segmentation such as graph cut segmentation weren't completely automatic. Feature extraction also highly depended on pixel labels that were semi-automatic. As the computational capabilities increased in the first part of the decade of the 21st century, algorithms moved towards complete automation and emergence of machine learning also began.

Phase. 2

This era began with automatic approaches to segment and classify using Machine learning based approach. Detection of Tuberculosis using CXR included a preprocessing step, in which images were resized, a segmentation step, in which region of interest is separated from the rest of Image, a feature extraction step, in which textural feature(opacity for tuberculosis in particular) and geometrical features(circular pattern) were extracted and finally a classification using machine learning based algorithms such as Support Vector Machine(SVM), Principal Component Analysis(PCA).

Pre-processing: Pre-processing is usually a first step done to resize all the images acquired to the same pixel size. Some approaches towards cropping and enhancement of images are also being attempted in various papers.

Segmentation: Separating area of interest from the rest of the image is called Segmentation. Various kinds of automatic techniques such as [14], [22] are developed. Efficiency and what kind of segmentation will work depends largely on the rest on the methodology which is used for feature extraction and classification too. Some of the popular techniques studied in this paper are as follows:

Graph-cut Segmentation: It is an old method of segmentation and semi-automatic version of graph cut segmentation has been in existence since early 2000s but completely automatic segmentation came a lot later with very accurate edge detection using M.L based methods [22]. Popularly it has two types of approaches, thresholding approach and edge specific approach.

Anatomical Atlases With Nonrigid Registration: In this approach the strategy [14]involves of three main steps, a content-based image retrieval approach for identifying training images, using SIFT-flow for deformable registration creating the initial patient-specific anatomical model of lung shape, and extracting refined lung boundaries using a graph cuts optimization approach.

Feature extraction: In image processing, features are a collection of information about the image. Properties of an image such as points, edges or objects can be called as features. Idea is to track circular opaque patterns since Tuberculosis bacteria acquire bubble-like patterns in pulmonary Tuberculosis.

Classification: Classification is a method of dividing images into Normal and Abnormal classes. Algorithms such as SVM and PCA have been popular for such use.

Phase. 3

Increased computational capabilities brought the tsunami of advancements to the area of neural networks and deep learning. Architecturally Neural networks are inspired from the Human brain[5].All the neurons of one layer are connected to all the neurons of a consecutive layer. Neural network with its unique ability to self discover patterns and features in an image has changed the paradigm. Focus of research today is on large scale implementation of algorithms of high accuracy, together with a new dimension of 3-D visualization with RGB full scale. Following methods using the Neural network are dominating research in current times.

Convolutional Neural Networks: Convolutional Neural network is a type algorithmic model which takes images as its input, assigns weights and biases, and based on training recognizes patterns[6]. Applications of Convolutional Neural Networks are in the field of Image Recognition, Object or Pattern Detection in an image, Image Classification e.t.c. Usually the architecture of CNN has a convolutional layer, pooling layer and a fully connected layer in different permutations and cinations according to application and computational capabilities. Function of the Convolution layer is to extract features from input images such as edges and colour contrast. Dot product of a known filter matrix is convolved with the pixel matrix of the input image to extract features. Pooling is connected next to the convolutional layer to decrease dimensions of the matrix after convolving. Pooling layer extracts dominant features from the image and also decreases the computational complexity. Finally a fully connected layer is used to flatten the matrix into a column vector on which different propagation and iterations are performed.

Modality-Specific Convolutional Neural Network: Modality specific deep learning CNN takes a unique approach to improve accuracy of common CNN[8]. Base CNN model is first trained on a collection of datasets of different diseases such as RSNA pneumonia, pediatric pneumonia and modality specific features are extracted and classified. Finally the trained model is cross validated on Images with pulmonary Tuberculosis.

Transferred Learning Based Deep Convolutional Neural Network: This approach smartly uses pre-trained convolutional neural networks that are publicly available, trained on very large image datasets and uses transferred learning to directly utilize the learned features and patterns[9][10][11].

Bayesian Convolutional Neural Network: Rare borderline case of Tuberculosis can be misdiagnosed using general CNN because it doesn't consider the problem of uncertainty in the softmax layer that classifies the normal or abnormal CXR[7]. The problem arises when probabilistic choices in the softmax layer aren't significantly different from each other. B-CNN uses calculations of variance in close cases to deal with the softmax interference issue. In this method, Variational inference which is calculated is used to reconfigure weights of CNN models.

Proceedings DOI: 10.21467/proceedings.114
Series: AIJR Proceedings
ISBN: 978-81-947843-8-8
ISSN: 2582-3922

3-D Vision Models and Visualization: 3-D Vision Models is a new approach, but not that extensively researched yet for Tuberculosis Detection because opaque pattern detection in CXR has good enough accuracy. But for borderline cases which are on borderline and diagnosis is difficult, this paradigm is being tested. Cross- sectional images from different techniques such as C.T scan and CXR are in research for this purpose[12]. Visualization is also being attempted to improve TB detection for borderline cases[13]. It is also imperative to note, visualization models are once again returning to segmentation methods in addition to deep learning models to explore even more as far as accuracy, sensitivity and specificity parameters are concerned for visualization[23].

2 Summary and Conclusion

Following table summarizes all the important papers studied for this paper. Table only mentions the best result acquired by respective papers. The review studies aspects of computer aided tuberculosis detection and concludes that deep learning convolutional methods are on par with human radiologists and will only improve with better datasets in future. Study also concludes Visualization methods with deep learning based segmentation and classification can help improve diagnosis of borderline cases.

Table I

Authors	**Journal/Conference, Year**	**Dataset**	**Results**	**Contribution**	**Research Gap**
Bram van et.al **[2]**	IEEE, 2002	**Database 1.** 326 abnormal, 290 normal and **Database.2**100 normal and 100 abnormal.	Average accuracy for whole database= 0.820 Maximum accuracy=0.986	Methods Of Segmentation and Classifications were explored using **Local Texture based analysis**	Low level of accuracy
Rui Shen et.al **[3]**	IEEE, 2010	**Database.1** 20 (PA) CXRs with TB cavities, 19 without TB cavities, and 17 normal cases **Database.2** contains 93 PA CXRs of normal cases.	True Positive rate = 82.35% False Positive rate= 0.435/image	Automated segmentation technique, which takes a **hybrid knowledge-based Bayesian classification approach to detect** TB cavities automatically was explored	Low Initial contour placement of TB infected area

Stefan Jaeger et.al [4]	IEEE 2014	**Database.1** 80 CXRs are normal and 58 CXRs are abnormal **Database.2** 340 normal CXRs and 275 abnormal CXRs with TB	AUC of 87% and 90%, respectively for the two datasets. Sensitivity of 74.1% and a specificity of 81.3%	Lung region using a graph cut segmentation method was explored.	Can be further optimized
Sema Candemir et.al [14]	IEEE 2014	**Dataset.1** JSRT. **Dataset.2** Montgomery County, MD, USA.	**Ω** JSRT = 0.954 MD =0.941 (ACD): JSRT = 1.321 MD = 1.216 Dice's Coeff: JSRT =0.967 MD = 0.96	A lung segmentation method based on non-rigid registration of pixels.	Can't distinguish between fluid filled lungs which are opaque.
R. H. H. M. Philipsen et.al [15]	IEEE, 2015	Six dataset of 100 CXR each with 50 normal and 50 abnormal images	JSRT = 0.90 Good ROC RESULTS	**Energy Based Normalization** is used to counter scanner settings. acquisition methods	Need of analysis on redundant features
Melendez et.al [24]	Nature , 2016	Normal =319 Abnormal = 73	Characteristic curve =0.84 specificity =95%	Ranking features to remove redundant features explored	Use of data from a single site.
R. Hooda et.al [18]	IEEE, 2017	Shenzen	Overall accuracy = 94.73% Validation accuracy = 82.09%	Test of 7 conv and 3 fully connected CNN	Transfer learning can be further tested
Imam Junaedi et.al [16]	IEEE, 2019	46 normal, 24ptb, 8 stb	Normal:62.82% accuracy, 86.96% sensitivity, 28.13% specificity PTB: 67.95% accuracy, 29.17% sensitivity, 85.19% specificity STB: 89.74% accuracy, 0%	**Gray Level Co-Occurrence Matrix (GLCM)** features as the input were explored. For Optimization PCA is used and for classification SVM is used.	Performance can be improved

ISBN: 978-81-947843-8-8
ISSN: 2582-3922

			sensitivity, 100% specificity		
P.Prasanna et.al [17]	IEEE, 2019	60% Abnormal, 40% Normal	Accuracy = 0.97 Sensitivity =0.89 Specificity = 0.99 (best of result in different approaches)	**Center based genetic algorithm** (CBGA) are used for Optimization, And Neural Network based classification	Neural Network architecture is vague
F. Pasa et.al [13]	Nature , 2019	Montgomery and Shenzhen datasets	Accuracy: MC = 79% SZ= 84.4% CB= 86.2% AUC: MC= 0.811 SZ=0.90 CB= 0.925	Attempt at fast diagnosis and Visualization	Pre-processing is limited to resizing of image
T. Rahman *et al.* [23]	IEEE, 2020	3500 TB infected CXR images, 3500 normal CXR	Accuracy = 96.47%, Precision = 96.62%, Sensitivity = 96.47%, F1-score = 96.47% and specificity = 96.51%	Transfer learning with ResNet18, ResNet50, ResNet101, ChexNet, InceptionV3, Vgg19, DenseNet201, SqueezeNet, and MobileNet explored	Further Segmentation methods can be explored.
Murphy et.al [25]	Nature , 2020	5565 Images with 15.3% TB Positive	98% specificity at 90% sensitivity	Commercial platform cost analysis along with good results	Cost analysis can be further explored

References

[1] WHO, Global TB report 2020

[2] Bram van Ginneken et .al, “Automatic Detection of Abnormalities in Chest Radiographs Using Local Texture Analysis.”,IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 21, NO. 2, FEBRUARY 2002

[3] Rui Shen et.al, “A Hybrid Knowledge-Guided Detection Technique for Screening of Infectious Pulmonary Tuberculosis From Chest Radiographs.”, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 57, NO. 11, NOVEMBER 2010

[4] Jaeger, S. et al. “Automatic tuberculosis screening using chest radiographs.” IEEE Trans. Med. Imaging 33, 233–245 (2014)

[5] S Haykin, Neural networks: A comprehensive foundation, Prentice Hall, 1998, ISBN 0-1327-3350-1.

[6] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[7] Z. Ul Abideen et al., "Uncertainty Assisted Robust Tuberculosis Identification With Bayesian Convolutional Neural Networks," in IEEE Access, vol. 8, pp. 22812-22825, 2020, doi: 10.1109/ACCESS.2020.2970023.

[8] S. Rajaraman and S. K. Antani, "Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs," in IEEE Access, vol. 8, pp. 27318-27326, 2020, doi: 10.1109/ACCESS.2020.2971257.

[9] Ho T.K.K., Gwak J., Prakash O., Song JI., Park C.M. (2019) Utilizing Pretrained Deep Learning Models for Automated Pulmonary Tuberculosis Detection Using Chest Radiography. In: Nguyen N., Gaol F., Hong TP., Trawiński B. (eds) Intelligent Information and Database Systems. ACIIDS 2019. Lecture Notes in Computer Science, vol 11432. Springer, Cham.

[10] T. Rahman et al., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," in IEEE Access, vol. 8, pp. 191586-191601, 2020, doi: 10.1109/ACCESS.2020.3031384.

[11] U.K. Lopes, J.F. Valiati,Pre-trained convolutional neural networks as feature extractors for tuberculosis detection,Computers in Biology and Medicine,Volume 89,2017,Pages 135-143, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2017.08.001.

[12] Fernando Llopis et.al., Using improved optical flow model to detect Tuberculosis, University of Alicante. Carretera San Vicente del Raspeig s/n 03690 San Vicente del Raspeig - Alicante

[13] Pasa, F., Golkov, V., Pfeiffer, F. *et al.* Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. *Sci Rep* 9, 6268 (2019). https://doi.org/10.1038/s41598-019-42557-4

[14] Candemir S., et.al, Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration, IEEE Trans. Med. Imaging, 33 (2) (2014), pp. 577-590

[15] Philipsen, R.H., Maduskar, P., Hogeweg, L., Melendez, J., Sánchez, C.I., van Ginneken, B.: Localized energy-based normalization of medical images: application to chest radiography. IEEE Trans. Med. Imaging 34(9), 1965–1975 (2015)

[16] Junaedi, Imam & Yudaningtyas, Erni & Rahmadwati, Rahmadwati. (2019). Tuberculosis Detection In Chest X-Ray Images Using Optimized Gray Level Co-Occurrence Matrix Features. 95-99. 10.1109/ICOIACT46704.2019.8938584.

[17] P. P. Kumari and B. P. Rao, "Automated Severity Analysis of TB Using Center Based Genetic Neural Network," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 349-353, doi: 10.1109/ICSSIT46314.2019.8987857.

[18] R. Hooda, S. Sofat, S. Kaur, A. Mittal and F. Meriaudeau, "Deep-learning: A potential method for tuberculosis detection using chest radiography," 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, 2017, pp. 497-502, doi: 10.1109/ICSIPA.2017.8120663.

[19] B. Van Ginneken, B. M. Ter Haar Romeny and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey," in IEEE Transactions on Medical Imaging, vol. 20, no. 12, pp. 1228-1241, Dec. 2001, doi: 10.1109/42.974918.

[20] K. Doi, M. Giger, R. Nishikawa, K. Hoffmann, H. MacMahon, R. Schmidt, et al., "Digital radiography: A useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images", *Acta Radiologica*, vol. 34, pp. 426-439, 1993.

[21] K. Abe, K. Doi, H. MacMahon, M. Giger, H. Jia, X. Chen, et al., "Computer-aided diagnosis in chest radiography preliminary experience", *Investigat. Radiol.*, vol. 28, no. 11, pp. 987-993, 1993

[22] R. Ramya and P. S. Babu, "Automatic tuberculosis screening using canny Edge detection method," 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 2015, pp. 282-285, doi: 10.1109/ECS.2015.7124909.

[23] T. Rahman et al., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," in IEEE Access, vol. 8, pp. 191586-191601, 2020, doi: 10.1109/ACCESS.2020.3031384.

[24] Melendez, J., Sánchez, C., Philipsen, R. *et al.* An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci Rep* 6, 25265 (2016). https://doi.org/10.1038/srep25265

[25] Murphy, K., Habib, S.S., Zaidi, S.M.A. *et al.* Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. *Sci Rep* 10, 5492 (2020). https://doi.org/10.1038/s41598-020-62148-y

ISBN: 978-81-947843-8-8 ISSN: 2582-3922