

Ensemble deep learning for tuberculosis detection

Mohd Hanafi Ahmad Hijazi¹, Leong Qi Yang², Rayner Alfred³, Hairulnizam Mahdin⁴, Razali Yaakob⁵

^{1,2}Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia

³Knowledge Technology Research Unit, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia

⁴Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Malaysia

⁵Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia

Article Info

Article history:

Received Apr 3, 2018

Revised Jun 4, 2018

Accepted Jun 18, 2018

Keywords:

Deep learning

Ensemble

Image classification

Medical image analysis

Tuberculosis detection

ABSTRACT

Tuberculosis (TB) is one of the deadliest infectious disease in the world. TB is caused by a type of tubercle bacillus called Mycobacterium Tuberculosis. Early detection of TB is pivotal to decrease the morbidity and mortality. TB is diagnosed by using the chest x-ray and a sputum test. Challenges for radiologists are to avoid confused and misdiagnose TB and lung cancer because they mimic each other. Semi-automated TB detection using machine learning found in the literature requires identification of objects of interest. The similarity of tissues, veins and small nodules presenting the image at the initial stage may hamper the detection. In this paper, an approach to detect TB, that does not require segmentation of objects of interest, based on ensemble deep learning, is presented. Evaluation on publicly available datasets show that the proposed approach produced a model that recorded the best accuracy, sensitivity and specificity of 91.0%, 89.6% and 90.7% respectively.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Mohd Hanafi Ahmad Hijazi,
Faculty of Computing and Informatics,
Universiti Malaysia Sabah,
Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.
Email: hanafi@ums.edu.my

1. INTRODUCTION

Tuberculosis (TB) is caused by the bacteria Mycobacterium tuberculosis, and it most often affects the lungs, it is a potentially serious infectious disease. Currently, in the global 2018 tuberculosis report, the World Health Organization (WHO) estimated that 1.6 million people being infected by TB died (1.2-1.4 million HIV-negative and 0.3 million HIV-positive) and 10.0 million had fallen ill [1]. Amongst HIV patients, TB is the leading cause of death [1].

Tuberculosis diagnosed mostly using the skin test, chest x-ray and a sputum test [2]. X-rays (CXR) can detect TB to a certain extent, but it cannot guarantee whether the patient has TB or some other infection due to difficulties in identifying malignant [3]. Nevertheless, CXR has become an important tool to detect TB due to the increased availability of radiography, in particular digital radiography, with better image quality and safety [4]. Early detection of lung related diseases such tuberculosis is pivotal to decrease the morbidity and mortality [5]. However, in Malaysia most lung cancer and tuberculosis cases are diagnosed late hence reduces chance of survival [6]. For century, challenges for radiologists to avoid confused and misdiagnose tuberculosis to other lung related diseases because they mimic each other [7]. A semi-automated system to effectively classify pulmonary nodules with low false positive rate is deemed necessary to assist radiologist to screen the chest radiograph images [8].

The application of machine learning and image processing approaches on medical images have been widely used, a comprehensive review is presented in [9] and some examples can be found in [10, 11]. Those work have produced superior classification performance compared to conventional approach. With

respect to semi-automated tuberculosis detection, examples of previous work have been reported in [9, 12, 13]. Two sets of features, in the form of object detection based and content based features, were extracted from the images with Support Vector Machine (SVM) was employed as classifier [9]. The best accuracy of 84.1% was recorded when applied on the Shenzhen dataset. Another work that applied image registration before object based features were extracted to detect tuberculosis recorded the best accuracy of 60% [12]. In [13], different types of features and analysis were employed to improve the classification result.

Recently, deep learning architectures have been employed to learn features from the medical images [14]. Deep learning is an approach that applies deep architectures that include many layers of non-linear information processing to learn features from data [15]. The deep learning techniques are categorized into two; deep discriminative network such as Convolution Neural Network (CNN) and unsupervised method such as deep belief network (DBN). Deep learning computes hierarchical features from images. The higher level features are defined from lower level features. Its ability to learn features at the higher level has been shown to produce better results. Examples of work that employs deep learning, in particular CNN, to detect tuberculosis can be found in [16] and [17]. Using a customized CNN produced the best accuracy of 82.09% [16]. In [17], an optimized CNN architecture was proposed where they manage to reduce the amount of computation without sacrificing the performance. The best accuracy reported was 79.0% and 84.4% for Montgomery and Shenzhen datasets respectively.

In machine learning domain, ensemble based classifiers are widely used to build predictive model by integrating multiple models [18]. There are a number of different ensemble techniques, the most used are bagging and boosting. The bagging method considers various outputs of learned classifiers into a single prediction by means of voting while the boosting approach repeatedly running a weak learner on different set of training data and then combines into a single strong classifier. To the best of our knowledge, there are limited number of previous work consider ensemble of deep learning architectures to learn features for TB detection, some examples can be found in [19-21]. However, the work presented limited at the pre-trained CNN (GoogLeNet, ResNet and VggNet) features only [19], where the best accuracy recorded was 84.7% on the Shenzhen dataset. The work presented in [20] employs AlexNet and GoogleNet with excessive augmentation (including radiologists' intervention) and recorded an Area Under Curve (AUC) of 99% using pre-trained models. An additional convolutional layer for feature extraction on top of the basic CNN was proposed in [21] whereby an accuracy of 84% was recorded for Shenzhen dataset.

Based on the literature, more work is required to achieve better TB detection rate, in its early stage in particular, with minimal human intervention. In this paper, an approach to ensemble different architectures of deep learning for tuberculosis detection is presented. The contribution of this paper is as follow:

- a) Three different deep learning architectures were employed to perform tuberculosis detection.
- b) Ensemble technique in the form of soft voting used to combine the different deep learning architectures employed in tuberculosis detection.

The proposed ensemble deep learning, in the form of CNN, for tuberculosis detection is presented in Section 2. Section 3 describes the experimental setup and discussion of results. Section 4 concludes this paper.

2. THE PROPOSED APPROACH

In this paper, we present an approach to detect TB using ensemble deep learning classifiers. Figure 1 shows the framework of TB detection using ensemble deep learning classifiers. It has three main modules: (i) image pre-processing, (ii) classifier generation, and (iii) ensemble classification. These modules were implemented across two stages; the training stage for deep learning model generation, and the test stage to perform TB detection. Details of each of the modules described above, with respect to the proposed approach, are presented in Sub-section 2.1 to 2.3 below.

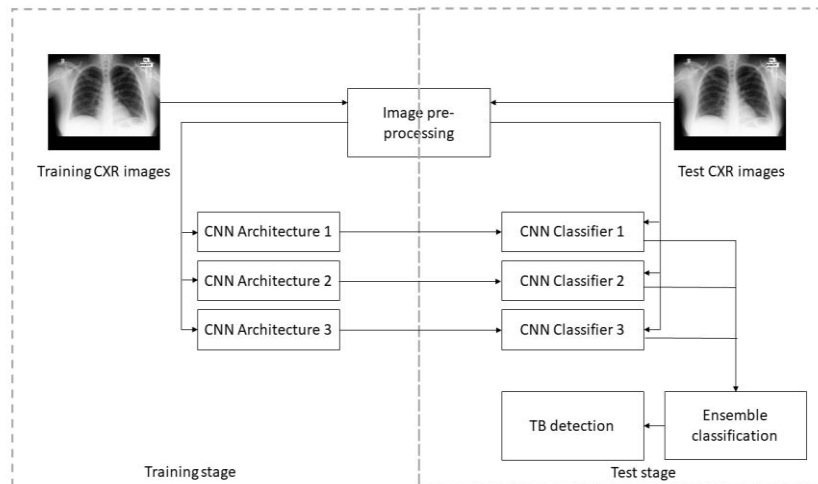


Figure 1. The proposed deep learning based TB detection approach

2.1. Image Pre-processing

The images acquired were first pre-processed to retain only the Region Of Interest (ROI). Figure 2 shows an example of CXR image. With respect to the work presented in this paper, the ROI is the lung area. To achieve this, a mask of the lung area was constructed for each CXR image. The constructed mask was then superimposed on the CXR image. Figure 3 shows an example of a masked image of Figure 2.



Figure 2. Chest X-Ray image

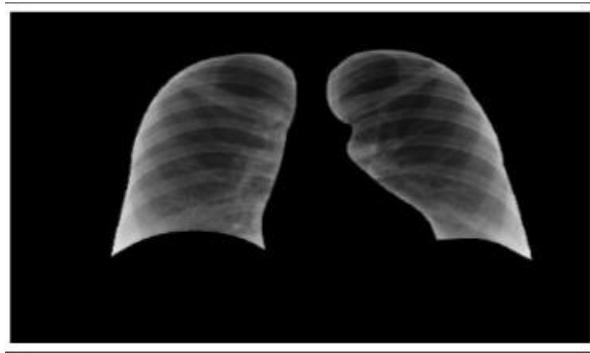


Figure 3. Chest X-Ray image with ROI identified

2.2. CNN Classifier Generation

In this paper, we employed three CNN architectures, InceptionV3, VGG16 and a custom-built architecture. CNN was selected due to its ability to extract and learn meaningful features on its own [8]. Further readings on InceptionV3 and VGG16 can be found in [22] and [23] respectively. With respect to the custom-built CNN architecture, we reduced the number of layers to 15. By reducing the number of layers, less number of parameters are generated for training and hence less time is required to complete the training. Each of the CNN architecture described above generated its own model based on the pre-trained model and the dataset used in this paper. The generated models were then used to perform TB detection.

2.3. Ensemble Classification

Each of the generated CNN model, as described in the foregoing sub-section, performed TB detection individually. With an aim to produce better detection result, ensemble classification was employed. With respect to the work presented in this paper, a majority voting was used. To predict the final label of each test image (TB or non-TB), the prediction made by individual CNN model was considered collectively and label with the highest vote was selected as the final label.

2.4. Performance Measurement

To measure the performance of the proposed approach, three measurement metrics were employed: sensitivity, specificity and accuracy. Sensitivity and specificity were used to measure how well the proposed approach correctly classify TB or non-TB cases. Accuracy was used to measure the overall performance of the proposed approach in detecting TB cases.

3. EXPERIMENTAL SETUP AND DISCUSSION OF RESULTS

The dataset used to evaluate the proposed work, the nature of the experiments conducted, and discussions of the results obtained are described in this section.

3.1. The Dataset

To evaluate the proposed approach presented in this paper, two public CXR images, Montgomery and Shenzhen datasets [24], are used. The Montgomery dataset consists of 80 normal images and 58 images with manifestations of TB. The size of the CXR is either 4020×4892 pixels or 4892×4020 pixels in Portable Network Graphics (PNG) format. The Shenzhen dataset consists of 662 images, where 336 are TB manifested CXRs. The size of the images are approximately 3000×3000 pixels.

3.2. Experimental Setup

The aim of the experiment conducted was to evaluate the performance of the proposed approach to TB detection using CXR images. To achieve this, two sets of experiments were conducted. The first was to evaluate the performance of using different CNN architectures to TB detection. The second was conducted to compare the performance of ensemble of classifiers used in experiment 1 with individual classifier. Ten-fold Cross Validation (TCV) was employed, whereby the dataset was randomly divided into ten equal sized subsets; and on each iteration one subset was used as the test set while the other was used as the training set. The images were resized to 300×300 pixels before they were fed to the CNN architectures for model generation. The size selected is similar to the work presented in [19] and [20]. With respect to the work presented in this paper, the Keras [25] implementation of InceptionV3 and VGG16 were used. Both architectures were pre-trained using Imagenet dataset.

3.3. Experimental Results

The results of the experiments described above are presented in Sub-section 3.3.1 and 3.3.2 respectively.

3.3.1 Performance of Different CNN Architectures to TB Detection

In this experiment, the images were classified as either TB or non-TB. For each CNN architecture, the learning rate and epoch were set to 1e-5 and 150 respectively. Sigmoid classifier was employed to train the model. Table 1 shows the results obtained. The best sensitivity, specificity and accuracy were recorded by the custom-built CNN architecture.

Table 1. The Performance of Different CNN Architectures to TB Detection

CNN Architecture	Sensitivity (%)	Specificity (%)	Accuracy (%)
InceptionV3	80.7	90.2	85.0
VGG16	81.7	90.3	85.6
Custom	88.6	91.7	90.0

3.3.2 Performance of Ensemble CNN to TB Detection

For the second set of experiment, the individual CNN classifiers were ensemble to perform the classification. Table 2 show the result obtained.

Table 2. The Performance of Ensemble CNN Architectures to TB detection

CNN Architecture	Sensitivity (%)	Specificity (%)	Accuracy (%)
Ensemble	89.6	90.7	91.0

3.4. Analysis of Results

Based on the results shown in Table 1 and 2, we observed that the proposed ensemble CNN produced an improved detection of TB cases (sensitivity) and overall accuracy by 1%. Although the specificity is decreased also by 1%, with respect to medical image analysis, sensitivity is deemed more pertinent as we do not want to misclassify positive case as negative. Further inspection on how the CNN classify the CXR into TB or non-TB, we found that the area that the network was focus on was both the upper lobes. Using the heat map overlay obtained from the fifth convolutional layer of InceptionV3 classifier, Figure 4 shows examples where the dark purple background represents area that were not activated while the dark red was heavily activated. For future work, it would be recommended to focus only on the activated area for TB detection.

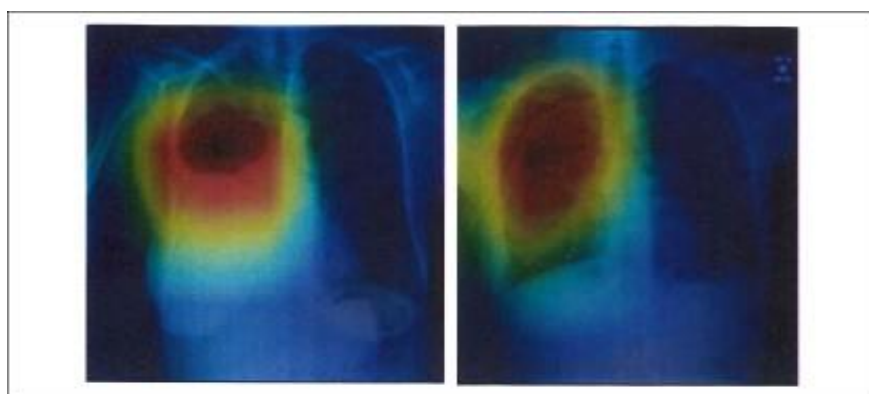


Figure 4. Activation heatmaps of CXRs using InceptionV3

4. CONCLUSION

This paper presented an approach to detect TB using ensemble deep learning based features and models. Three different CNN architectures were employed, InceptionV3, VGG16 and custom-built architecture. The results obtained shows that, the ensemble of the three classifiers using majority voting, produced the best TB detection performance. Further inspection using InceptionV3 architecture, it is found that the upper lobes of the CXR were the most activated area which indicates the symptoms of TB could be found mostly in that area. Hence, future work may focus the identification of features only on those areas which could improve the detection performance with respect to accuracy and reducing processing time.

ACKNOWLEDGEMENTS

This work was funded by Universiti Malaysia Sabah (UMS) through a grant SDK0018-2017 and Artificial Intelligence Research Unit, UMS.

REFERENCES

- [1] N. Fogel, "Tuberculosis: A disease without boundaries," *Tuberculosis*, vol. 95(2015), pp. 527-531, 2015.
- [2] M. Sabait, *et al.*, "Challenges and opportunities to improve tuberculosis screening among immigrant plantation workers in Sabah, Malaysia," *International Journal of Public Health*, vol. 5(1), pp. 41-45, 2016.
- [3] M. Nasr and A. A. A. El-Mageed, "Using image mining techniques for optimizing the treatment methods of lung cancer," *Journal of Multidisciplinary Engineering Science and Technology*, vol. 3(1), pp. 3613-3620, 2016.
- [4] World Health Organization, "Chest radiography in tuberculosis detection," *World Health Organization*, 2016.
- [5] A. O. Ankrah, *et al.*, "Tuberculosis," *Seminars in Nuclear Medicines*, vol. 48(2), pp. 108-130, 2018.
- [6] K. C. Siang and C. K. M. John, "A review of lung cancer research in Malaysia," *Med J Malaysia*, vol. 71, pp. 70-78, 2016.
- [7] I. Hammen, "Tuberculosis mimicking lung cancer," *Respir Med Case Reports*, 16, pp. 45-47, 2015.
- [8] S. S. Meraj, *et al.*, "Artificial intelligence in diagnosing tuberculosis: A review," *International Journal on Advanced Science Engineering Information Technology*, vol. 9(2019), pp. 81-91, 2019.
- [9] C. S. Poornimadevi and C. Helen Sulochana, "Automatic detection of pulmonary tuberculosis using image processing techniques," *IEEE WiSPNET 2016*, 2016, pp. 798-802.
- [10] M. A. Kahya, "Classification enhancement of breast cancer histopathological image using penalized logistic regression," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13(1), pp. 405-410, 2019.

- [11] Y. Zheng, *et al.*, "Automated disease/no disease grading of age-related macular degeneration by an image mining approach," *Investigative Ophthalmology and Visual Science*, vol. 53(13), pp. 8310-8318, 2012.
- [12] S. Jaeger, *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans Med Imaging*, vol. 33(2), pp. 233-245, 2014.
- [13] L. Hogeweg, *et al.*, "Automatic detection of tuberculosis in chest radiographs using a combination of textural, focal, and shape abnormality analysis," *IEEE trans Med Imaging*, vol. 34(12), pp. 2429-2442, 2015.
- [14] G. Litjens, *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, 2017.
- [15] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends in Signal Processing*, vol. 7(3-4), pp. 197-387, 2014.
- [16] R. Hooda, *et al.*, "Deep-learning: A potential method for tuberculosis detection using chest radiography," in *Proc. Of the 2017 IEEE International Conference on Signal and Image Processing Applications*, 2017, pp. 497-502.
- [17] F. Pasa, *et al.*, "Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization," *Scientific Reports*, vol. 9(6268), 2019.
- [18] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33(1-2), pp. 1-39, 2010.
- [19] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Computers in Biology and Medicine*, vol. 89(2017), pp. 135-143, 2017.
- [20] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284(2), pp. 574-582, 2017.
- [21] S. Hwang, *et al.*, "A novel approach for tuberculosis screening based on deep convolutional neural networks," In *Proceedings of SPIE: Medical Imaging*, Vol 9785, 2016.
- [22] C. Szegedy, *et al.*, "Rethinking the inception architecture for computer vision," In *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.
- [24] S. Jaeger, *et al.*, "Two Public Chest X-Ray Datasets for Computer-Aided Screening of Pulmonary Diseases," *Quant Imaging Med Surg*, vol. 4(6), pp. 475-477, 2014.
- [25] World Health Organization, "Global tuberculosis report 2018," *World Health Organization*, 2018.

BIOGRAPHIES OF AUTHORS



Mohd Hanafi Ahmad Hijazi is an Associate Professor of Computer Science at the Faculty of Computing and Informatics, Universiti Malaysia Sabah in Malaysia. His research work addresses the challenges in knowledge discovery and data mining to identify patterns for prediction on structured and/ or unstructured data; his particular application domains are medical image analysis and understanding and sentiment analysis on social media data. He has authored/co-authored more than 40 journals/ book chapters and conference papers, most of which are indexed by Scopus and ISI Web of Science. He also served on the program and organizing committees of numerous national and international conferences.



Leong Qi Yang is a graduate of Software Engineering of Universiti Malaysia Sabah. He is currently working as a software developer at Asuka Thai, where he helps to turn the legacy manufacturing software into cloud applications.



Rayner Alfred is an Associate Professor of Computer Science at the Faculty of Computing and Informatics, Universiti Malaysia Sabah in Malaysia. He leads and defines projects around knowledge discovery and information retrieval that focuses on building smarter mechanism that enables knowledge discovery in structured and unstructured data. His work addresses the challenges related to big data problem. He has authored and co-authored more than 85 journals/book chapters, conference papers and editorials.



Hairulnizam Mahdin is an Associate Professor at Department of Information Security and Web, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia. He completed his Ph.D. thesis in 2012 and was conferred a doctorate degree from Deakin University Australia in the same year. His research focuses on the area of data management, IoT, RFID, information security, software engineering and web technology. He has published his research in both ISI and Scopus-indexed journals.



Assoc. Prof. Razali Yaakob received the Bachelor Degree in Computer Science in 1996 and Master in Computer Science from in 1999, from Universiti Putra Malaysia, and PhD from University of Nottingham, United Kingdom in 2008. Currently, he is a lecturer at the Faculty of Computer Science and IT, Universiti Putra Malaysia. His research areas include artificial neural network, pattern recognition, and evolutionary computation in game playing. He is a member of the Intelligent Computing Group at the faculty.