Special Review

Artificial Intelligence, Radiology, and Tuberculosis: A Review

Sagar Kulkarni, MBBS, BSc, Saurabh Jha, MBBS, MRCS, MS

Abbreviations

TB Tuberculosis

AI Artificial Intelligence

HIV Human Immunodeficiency Virus

AIDS Acquired Immune Deficiency Syndrome

CAD Computer-Aided Diagnosis

KIT Korean Institute of Tuberculosis

NIH National Institutes of Health

AUC Area Under the Receiver-Operator Curve

FLOP Floating Point Operations

Tuberculosis is a leading cause of death from infectious disease worldwide, and is an epidemic in many developing nations. Countries where the disease is common also tend to have poor access to medical care, including diagnostic tests. Recent advancements in artificial intelligence may help to bridge this gap. In this article, we review the applications of artificial intelligence in the diagnosis of tuberculosis using chest radiography, covering simple computer-aided diagnosis systems to more advanced deep learning algorithms. In so doing, we will demonstrate an area where artificial intelligence could make a substantial contribution to global health through improved diagnosis in the future.

Key Words: Artificial intelligence; Deep learning; Computer-aided diagnosis; Tuberculosis; Global health.

INTRODUCTION

Tuberculosis (TB) is an epidemic in many parts of the world, being responsible for 1.6 million deaths in 2017; in the same year, 10 million people developed the disease (1). TB is the leading cause of death from infectious disease worldwide and disproportionately affects developing regions, such as Africa and South-East Asia (1). Frequently, countries that suffer from a high TB prevalence are also resource-poor, lacking in medical staff and equipment.

Perhaps the most pressing issue relating to TB is the emergence of multidrug resistant strains of the disease. About 558,000 people developed TB resistant to the most common anti-TB drug, rifampicin, in 2017; 82% of these patients had multidrug resistant TB (1). Identification of such patients and controlling resistant strains of the disease will be a significant challenge in the developing world for decades to come.

Artificial intelligence (AI) has been posited as a solution to assist in the fight against TB. AI's applications in diagnostic radiology may be able to provide accurate means of detecting the disease for low-income nations. In this article, we will review the literature on AI and TB imaging, suggesting future avenues for development and innovation.

CHEST RADIOGRAPHY AND TB

The radiographic presentation of pulmonary TB is varied, making it a challenging diagnosis. Firstly, it is important to distinguish between active and latent TB. Active TB is characterized by the presence of consolidation and cavitary lesions in the lungs, and has a high risk of infectious spread (2). By contrast, latent TB is characterized by stable fibronodular changes, such as scarring and nodular opacification, and has a

Acad Radiol 2019; ■:1–5

From the Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA. Received July 10, 2019; revised September 22, 2019; accepted October 5, 2019. Address correspondence to: S.K. e-mail: Sagar.Kulkarni@pennmedicine.upenn.edu

https://doi.org/10.1016/j.acra.2019.10.003

low risk of infectious spread (although the disease can be reactivated later in life) (2). Latent TB can also present with no radiographic signs. In practice, there is considerable overlap between the appearance of active and latent TB; the only definitive method to distinguish them is by observing for temporal resolution over a period of 4−6 months, where radiographic stability indicates the presence of latent TB. Both active and latent TB requires medical treatment, however, from a public health standpoint, active TB is more concerning due to its infective risk.

Active TB itself can be subdivided into different forms: primary TB, post-primary TB, and miliary TB. Primary TB demonstrates patchy consolidation on chest radiography, which can be present anywhere in the lung fields. The condition is frequently associated with a pleural effusion. Post-primary TB presents similarly, but is more likely to present with cavitations and shows a predilection for the apical and upper lung zones. Miliary TB shows multiple, diffusely located nodules in both lung fields and carries are poor prognosis (2). All of these descriptions constitute active TB, demonstrating the challenge in screening for the condition using chest radiography.

In patients with human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS), the appearance of TB varies depending on the severity of the disease. In a screening examination, AIDS patients with T-cell CD4+ counts above 200 cells/mm^3 exhibit signs of typical post-primary TB; conversely, when the CD4+ count falls to below 200 cells/mm^3, atypical signs, such as lower zone opacification, pleural effusion, or mediastinal adenopathy are more common (3). Since AIDS and TB are frequently comorbid in the developing world, the features of TB in these patients are particularly pertinent.

Furthermore, when reading a chest radiograph, other diagnoses must be considered. The differential diagnosis for consolidation is broad, including different forms of pneumonia, neoplastic disease, edema, and hemorrhagic disease. Therefore, not only is TB a challenging diagnosis to make due to its wide-ranging presentation, but it also has several mimics that have a similar radiological appearance. The combination of these factors makes the interpretation of TB screening examinations difficult for untrained observers.

The diagnosis of TB on a chest radiograph varies considerably between interpreting physicians. Linh et al. conducted a case-control study on TB chest radiographs, which demonstrated only a moderate interrater reliability (weighted kappa coefficient) of 0.60 and 0.67 for two different classification systems (4). Furthermore, the diagnosis of TB has relatively poor overall accuracy, with Kumar et al. (5) reporting a sensitivity and specificity of 78% and 51% respectively in a study of 75 radiographs from a Nepalese center; the accuracy is better for more distinctive findings, such as miliary TB, with another study reporting a sensitivity of 59%−69% and specificity of 97%−100% for this finding (6).

For patients with HIV/AIDS, the figures are similarly poor. In a study of 403 HIV-positive children sited in Burkina Faso, Cambodia, Cameroon and Vietnam, the interrater agreement (kappa coefficient) was between 0.16 and 0.36 for a local radiologist, a pediatric radiologist, and a pediatric pulmonologist. Chest radiography had a sensitivity of 71.4% and specificity of 50.0% for the diagnosis of TB (7).

Due to the rising need for radiological expertise in the developing world, investigators have developed software programs to detect TB on chest radiographs. Investigators faced the challenge of developing tools that could detect the wide range of possible pulmonary findings in active TB, including primary, post-primary, and miliary disease. Early attempts at this task used computer-aided diagnosis (CAD), akin to the systems currently used in mammography.

COMPUTER-AIDED DIAGNOSIS

In its early development, TB detection using AI was studied with conventional CAD (8). In CAD systems, detection is performed using a manually created preset feature model and the computer does not "learn" given a greater caseload. As a result, it has proved difficult to incorporate the varied presentation of TB into a single CAD system. CAD systems developed from 1996 to 2013 achieved accuracies ranging from 42% to 100%, usually for specific features such as the presence of cavitations (8). Also, the datasets used were often small, typically numbering in the 100s (8), limiting the applicability of CAD systems in a global health situation. What follows is a selection of CAD algorithms to illustrate its development and use.

Over the course of the past 3 decades, investigators developed numerous CAD algorithms for TB detection using varying methodologies. In 2013, Xu et al. (9) developed a CAD algorithm for TB cavity detection; the program worked by using classification tools to select regions of interest on chest radiographs that had features resembling TB cavities. The investigators used a small dataset of 35 chest radiographs containing 50 cavities, all obtained from The University of Alberta Hospital (Edmonton, Canada), and achieved a sensitivity of 78.8%, specificity of 86.8%, and overall accuracy of 82.8%. Though the program was accurate, its functional restriction to the detection of cavities, when, in reality, the radiographic presentation of TB is diverse and varied, is a limitation to its introduction to real-world clinical situations.

Using different methods, Song et al. (10) developed a CAD program to detect focal TB. In this case, a feature model was used to detect and computationally extract the ribs from the image. Without the ribs, the lung fields should be entirely radiolucent, thereby making any focal opacity obvious. The opacities were identified using morphological and region-growing operations, achieving an accuracy of 85% based on an original dataset of 200 images.

CAD programs have also been developed to detect a greater range of TB features. Using a combination of several imaging manipulation techniques, including masking and texture analysis, Jaeger et al. (11) developed a TB detection pipeline that analyzed chest radiographs in a step-wise fashion to suggest a diagnosis of TB. The authors used a dataset of 138 chest radiographs encompassing the varied presentation

of primary TB, rather than focusing solely on cavitary or focal lesions. An accuracy of 83% was achieved, which approaches human performance for the task of detecting primary TB.

Advancement has continued further in the field of CAD. Adding to the complexity of previous work, in 2018 Vajda et al. (12) developed a TB detection program that worked by first segmenting the lung fields, then extracting selected features in the image, such as certain shapes, which are analyzed by the classifier to determine if TB is present. The algorithm achieved an area under the receiver-operator curve (AUC) of 0.99 and an accuracy of 95.6% on the Shenzhen dataset, which was also used in multiple studies involving deep learning models, as detailed in the *Deep Learning* section of this article.

The only commercially available CAD-based TB detection software is CAD4TB (Delft Imaging Systems, Veenendaal, The Netherlands). The diagnostic performance of CAD4TB was evaluated in a systematic review by Pande et al. (13), which included 5 studies that tested the algorithm's performance in patients from Zambia, Tanzania, South Africa, and the United Kingdom. The systematic review found that the algorithm's AUC ranged from 0.71 to 0.84, indicating that commercially available products are lagging behind the forefront of AI development. Nevertheless, CAD4TB has been studied as a screening tool in Bangladesh (14) and Pakistan (15).

CAD4TB has also been used in conjunction with clinical information. Melendez et al. (16) combined the CAD4TB algorithm score, which is based on imaging findings, with 12 clinical features, such as the presence of hemoptysis, night sweats or an elevated axillary temperature. The augmented CAD4TB algorithm produced an AUC of 0.84, with a sensitivity of 49% and a specificity of 95%. Therefore, the use of clinical features led to little improvement in the accuracy of the algorithm.

Intriguingly, CAD4TB has now been redeveloped as a deep learning model rather than a CAD program. The latest version of the model, released in 2019, was trained on cohort of 500 labelled images from Pakistan and achieved a sensitivity of 90% and specificity of 98% upon testing for the task of detecting TB, surpassing all previous version of the algorithm (17).

DEEP LEARNING

Recently, AI deep learning networks have been developed for TB detection. The first of these was developed by Hwang et al. in 2016 (18). The authors used a model called AlexNet, a pretrained deep learning network that had previously achieved success in the ImageNet Large Scale Visual Recognition Competition—an image classification challenge for nonmedical images. AlexNet had already been pretrained for image recognition—this capability merely needed to be adjusted for use in medical imaging, as opposed to training the model from scratch. This concept in known as transfer learning in computer science and it allows investigators to use smaller training datasets for deep learning applications.

The authors used a dataset of 10,848 chest x-rays supplied by the Korean Institute of Tuberculosis (KIT), of which 70% were used to train the algorithm. For testing, 15% of the images from the KIT dataset were used in conjunction with 138 images from the National Institutes of Health (NIH), USA, and 662 images from Shenzhen No 3 People's Hospital, China, to demonstrate the cross-dataset validity of the model. An AUC of 0.964 was achieved using the KIT dataset, and AUCs of 0.88 and 0.93 were achieved for the NIH and Shenzhen datasets, respectively.

Overall, the model showed good screening performance in all three datasets, thereby illuminating a future where TB screening could be accomplished using AI and chest radiography. However, differing AUC performance between the datasets demonstrates a primary weakness of AI—overfitting, where a model can perform well on images from the same source dataset that it was trained on, even if the images themselves are unique to the model; an example of this would be images taken in the same hospital. Conversely, the model may perform less well on images from other datasets, such as images from a different country, even if the subject (such as TB) is the same. Such differences may be the result of varying image acquisition parameters, radiographic appearance, diagnostic criteria, and labelling practices across the globe, or even between different hospitals in the same city.

AI TB detection was improved upon by Lakhani and Sundaram (19). The authors used a dataset of 1,007 images, with 68% being used to train the algorithms. Pretrained and untrained models of GoogLeNet and AlexNet, two popular deep learning networks that achieved success in the ImageNet challenge, were tested. The effect of transfer learning was clear—untrained versions of AlexNet and GoogLeNet received AUCs of 0.90 and 0.88, whereas the pretrained versions achieved AUCs of 0.98 and 0.97, respectively. The most accurate approach utilized an ensemble of both models together with a radiologist to adjudicate discrepant cases, which achieved a sensitivity of 97.3%, specificity of 100%, and AUC of 0.99. The images used in the study were sourced from four different datasets, thereby allowing the model to demonstrate cross-dataset validity, an important characteristic given the potential global applicability of the technology.

AI can also assist in performing feats that have not been accomplished by human visual perception alone. Jaeger et al. (20) developed a deep learning model with the aim of classifying drug-resistant and drug-sensitive TB directly from the radiographic appearance. The authors used a relatively small image dataset of 135 images, of which 45% were TB sensitive cases and 54% were multidrug resistant cases. The model achieved an AUC of 0.66. While the performance of the model does not meet the expectations for a clinically applicable tool, since a random coin toss would achieve an AUC of 0.5, the results of the study illuminate an avenue of AI research that may become fruitful in the future. Indeed, the potential of AI to provide complementary information beyond the comprehension of humans is currently being actively researched for other applications.

TOWARD REFINEMENT IN DEEP LEARNING

Most recently, models for TB detection are becoming more refined and streamlined. Previously described deep learning models, such as AlexNet and GoogLeNet, are pretrained on millions of images and are capable of distinguishing between thousands of classes of images even before they are used in the context of radiology. As a consequence, they require substantial computer memory and hardware requirements to function effectively, even when working on the narrow task of detecting TB on a chest radiograph (21).

In light of this, in 2019 Pasa et al. (21) developed a deep learning model that was trained from scratch on chest radiographs only, for the sole task of TB detection. The authors trained the algorithm on the NIH and Shenzhen datasets (22) (the same datasets used by Hwang et al. (18), Lakhani and Sundaram, (19) and Vajda et al. (12)), as well as the Belarus Tuberculosis Portal dataset (23) (used by Lakhani and Sundaram (19)), totaling 1104 images. The model achieved an AUC of 0.811 on the NIH dataset, 0.9 on the Shenzhen dataset and 0.925 on a combined dataset, making the accuracy comparable with Hwang et al., (18) but inferior to Vajda et al. (12) CAD algorithm and Lakhani and Sundaram's (19) ensemble of AlexNet and GoogLeNet.

The strength of the Pasa et al. (21) model lies in its simplicity. While their network only includes 230,000 parameters, GoogLeNet includes 7 million parameters and AlexNet includes 60 million parameters. Alternatively stated, comparing the computational performance requirements of the respective models, the Pasa et al. model has 350 mega-floating point operations (mega-FLOPs), AlexNet has 1.5 giga-FLOPs and GoogLeNet has 3 giga-FLOPs. The advantage of using a simpler model is that it can be run on a cheaper processor. Pasa et al. noted that the processor used in their study cost less than US$200. Lower memory and hardware requirements are critical to the deployment of deep learning models in real clinical situations, particularly in resource-poor countries where the need is greatest.

CONCLUSION

The use of AI to detect TB in chest radiographs has progressed significantly in the past 3 decades. The field began with the development of CAD programs, which showed promising results for limited narrow applications. However, these tools were limited by their small dataset size and their reliance on preset feature models, a fundamental weakness of the technology itself. In the last 5 years, deep learning has made rapid progress in medical imaging, with TB detection emerging as an area of research interest. The new deep learning models have the potential to surpass the accuracy of their CAD predecessors and, therefore, could be used in a clinical setting in areas the world where TB is endemic. As Lakhani and Sundaram (19) suggested, the best use of such algorithms may be to augment to capabilities of radiologists working in resource-poor regions.

REFERENCES

1. World Health Organization. Global Tuberculosis Report 2018. 2018.
2. Nachiappan AC, Rahbar K, Shi X, et al. Pulmonary tuberculosis: role of radiology in diagnosis and management. RadioGraphics 2017; 37(1):52–72. [cited 2019 Jul 6] Available from: http://pubs.rsna.org/doi/10.1148/rg.2017160032.
3. Keiper MD, Beumont M, Elshami A, et al. CD4 T lymphocyte count and the radiographic presentation of pulmonary tuberculosis. Chest 1995; 107(1):74–80. [cited 2019 Jul 6] Available from: http://www.ncbi.nlm.nih.gov/pubmed/7813316.
4. Linh NN, Marks GB, Crawford ABH. Radiographic predictors of subsequent reactivation of tuberculosis. Int J Tuberc Lung Dis 2007; 11 (10):1136–1142. [cited 2019 Jul 9] Available from: http://www.ncbi.nlm.nih.gov/pubmed/17945072.
5. Kumar N, Bhargava SK, Agrawal CS, et al. Chest radiographs and their reliability in the diagnosis of tuberculosis. JNMA J Nepal Med Assoc 2005; 44(160):138–142. [cited 2019 Jul 9] Available from: http://www.ncbi.nlm.nih.gov/pubmed/16751817.
6. Kwong JS, Carignan S, Kang E-Y, et al. Miliary tuberculosis. Chest 1996; 110(2):339–342. [cited 2019 Jul 9] Available from: http://www.ncbi.nlm.nih.gov/pubmed/8697830.
7. Berteloot L, Marcy O, Nguyen B, et al. Value of chest X-ray in TB diagnosis in HIV-infected children living in resource-limited countries: the ANRS 12229-PAANTHER 01 study. Int J Tuberc Lung Dis 2018; 22(8):844–850. [cited 2019 Jul 9] Available from: http://www.ncbi.nlm.nih.gov/pubmed/29991391.
8. Jaeger S, Karargyris A, Candemir S, et al. Automatic screening for tuberculosis in chest radiographs: a survey. Quant Imaging Med Surg 2013; 3 (2):89–99. [cited 2019 May 19] Available from: http://www.ncbi.nlm.nih.gov/pubmed/23630656.
9. Xu T, Cheng I, Long R, et al. Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs. EURASIP J Image Video Process 2013; 2013(1):3. [cited 2019 Jul 6] Available from: https://jivp-eurasipjournals.springeropen.com/articles/10.1186/1687-5281-2013-3.
10. Song Y-L, Yang Y. Localization algorithm and implementation for focal of pulmonary tuberculosis chest image. In: 2010 International Conference on Machine Vision and Human-machine Interface, IEEE; 2010:361–364 http://ieeexplore.ieee.org/document/5532734/.
11. Jaeger S, Karargyris A, Antani S, et al. Detecting tuberculosis in radiographs using combined lung masks. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE; 2012:4978–4981 http://www.ncbi.nlm.nih.gov/pubmed/23367045.
12. Vajda S, Karargyris A, Jaeger S, et al. Feature Selection for Automatic Tuberculosis Screening in Frontal Chest Radiographs. J Med Syst 2018; 42(8):146. [cited 2019 Sep 22] Available from: http://www.ncbi.nlm.nih.gov/pubmed/29959539.
13. Pande T, Cohen C, Pai M, Ahmad Khan F. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. Int J Tuberc Lung Dis 2016; 20(9):1226–1230. [cited 2019 Sep 21] Available from: http://openurl.ingenta.com/content/xref?genre=article&issn=1027-3719&volume=20&issue=9&spage=1226.
14. Rahman MT, Codlin AJ, Rahman MM, et al. An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients. Eur Respir J 2017; 49(5):1602159. [cited 2019 Sep 21] Available from: http://www.ncbi.nlm.nih.gov/pubmed/28529202.
15. Zaidi SMA, Habib SS, Van Ginneken B, et al. Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on Chest radiography among private sector patients in Pakistan. Sci Rep 2018; 8 (1):12339. [cited 2019 Sep 21] Available from: http://www.nature.com/articles/s41598-018-30810-1.
16. Melendez J, Sánchez CI, Philipsen RHHM, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. Sci Rep 2016; 6:25265. [cited 2019 Sep 22] Available from: http://www.ncbi.nlm.nih.gov/pubmed/27126741.
17. Murphy K, Habib SS, Zaidi SMA, et al. Computer aided detection of tuberculosis on chest radiographs: an evaluation of the CAD4TB v6 system. 2019[cited 2019 Sep 21]; Available from: http://arxiv.org/abs/1903.03349
18. Hwang S, Kim H-E, Jeong J, et al. A novel approach for tuberculosis screening based on deep convolutional neural networks. In: Tourassi GD, Armato SG, eds. Proceedings of SPIE; 2016. p. 97852W.

http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2216198.

19. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017; 284(2):574–582. [cited 2019 May 9] Available from: http://pubs.rsna.org/doi/10.1148/radiol.2017162326.
20. Jaeger S, Juarez-Espinosa OH, Candemir S, et al. Detecting drug-resistant tuberculosis in chest radiographs. Int J Comput Assist Radiol Surg 2018; 13(12):1915. [cited 2019 May 17] Available from: http://www.ncbi.nlm.nih.gov/pubmed/30284153.
21. Pasa F, Golkov V, Pfeiffer F, et al. Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization. Sci Rep 2019; 9:6268. [cited 2019 Sep 21] Available from: http://www.nature.com/articles/s41598-019-42557-4.
22. Jaeger S, Candemir S, Antani S, et al. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg 2014; 4(6):475–477. [cited 2019 Sep 22] Available from: http://www.ncbi.nlm.nih.gov/pubmed/25525580.
23. Belarus tuberculosis portal. [cited 2019 Sep 22]. Available from:http://tuberculosis.by/.