

A Fully Automatic AI-based CT Image Analysis System for Accurate Detection, Diagnosis, and Quantitative Severity Evaluation of Pulmonary Tuberculosis

Chenggong Yan

Southern Medical University Nanfang Hospital

Lingfeng Wang

School of Information and Communication Engineering, University

Jie Lin

Southern Medical University Nanfang Hospital

Jun Xu

Southern Medical University Nanfang Hospital

Tianjing Zhang

Philips Medical Systems: Philips Healthcare

Jin Qi

University of Electronic Science and Technology of China

Xiangying Li

Xiangya Hospital Central South University

Wei Ni

Philips Medical Systems: Philips Healthcare

Guangyao Wu

Maastricht University Hospital: Maastricht Universitair Medisch Centrum+

Jianbin Huang

Southern Medical University Nanfang Hospital

Yikai Xu (✉ yikaixu917@gmail.com)

Southern Medical University Nanfang Hospital

Henry C. Woodruff

Maastricht University: Universiteit Maastricht

Philippe Lambin

Maastricht University: Universiteit Maastricht

Research

Keywords: Pulmonary tuberculosis, Artificial intelligence, Deep learning, Computed tomography, Image analysis

Posted Date: May 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-434810/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Accurate and rapid diagnosis of pulmonary tuberculosis (TB) plays a crucial role in timely prevention and appropriate medical treatment to the disease. This study aims to develop and evaluate an artificial intelligence (AI)-based fully automated CT image analysis system for detection, diagnosis, and burden quantification of pulmonary TB.

Methods: From December 2007 to September 2020, 892 chest CT scans from pathogen-confirmed TB patients were retrospectively included. A deep learning based cascading framework was connected to create a processing pipeline. To train and validate the model, 1921 lesions were manually labeled, classified by six categories of critical imaging features, and visually scored for the lesion involvement as the ground truth. “TB score” was calculated by the network-activation map to assess the disease burden quantitatively. Independent test datasets from two additional hospitals and NIH TB Portal were used to validate externally the performance of the AI model.

Results: CT scans from 526 participants (mean age, 48.5 years \pm 16.5; 206 women) were analyzed. The lung lesion detection subsystem yielded a mean average precision of 0.68 on the validation cohort. In the independent datasets, the overall classification accuracy for six pulmonary critical imaging findings indicative of TB were 81.08%-91.05%. A moderate to strong correlation was demonstrated between the AI model quantified “TB score” and the radiologist-estimated CT score.

Conclusion: This end-to-end AI system based on chest CT can achieve human-level diagnostic performance, and holds great potential for early management and medical resource optimization of patients with pulmonary TB in clinical practice.

Background

Tuberculosis (TB), caused by the bacillus Mycobacterium, has led to huge public health problem due to its rapid spread [1]. It is the leading cause of death by infectious disease worldwide [2]. Although the global morbidity and mortality of TB have been slowly declined, the disease burden remains substantial in endemic countries [3]. Chest imaging plays a crucial role in the work-up of patients with pulmonary TB [4]. In particular, computed tomography (CT) has been used to help diagnose, monitor imaging changes and evaluate the disease severity of pulmonary TB [5]. On the other hand, the activity and progressiveness of TB can be suggestive of radiographic features on CT [6]. This will support decision making of clinicians for timely isolation and appropriate treatment. However, the complexity of chest CT images have caused challenging workloads to the specialists. In countries where health and surveillance systems are weak, underdiagnosis and missed-reporting of TB is common.

Artificial intelligence (AI) has gained significant attention in recent years and many applications have been proposed in medical image recognition and interpretation [7]. Deep learning (DL), as the core technique of the rising AI, has made great progress in medical image analysis including skin diseases classification [8], diabetic retinopathy detection [9], lung cancer screening in chest CT [10]. Recent promising advances in the CT-based deep learning system demonstrated the potential of AI-assisted radiological diagnostic technology [11–13]. For example, a few studies have already reported supervised deep learning to diagnose, quantify extent, and predict triage of Coronavirus disease 2019 (COVID-19) disease on CT during the global pandemic [14]. Therefore, benefiting from end-to-end deep learning algorithms, we hypothesized that DL networks can be designed and established by automatic and adaptive learning the features through backpropagation, to achieve human expert-level performance for diagnosis and follow-up.

In this study, we developed and evaluated an AI based fully automated CT image analysis model to provide support in the detection, diagnosis, and disease severity quantification for patients with pulmonary TB.

Methods

Study participants

The study was approved by our institutional review board. Written informed consent was waived for this retrospective study. Between Dec, 2017 to Sep, 2020, we searched chest CT scans of patients with suspected TB from the picture archiving and communication system. A total of 1356 CT scans from 865 patients were collected who met the following inclusion criteria (Fig. 1): (a) age ≥ 18 years; (b) CT imaging examinations for known or suspected to have primary or secondary TB; (c) CT slice thickness $< 1.5\text{mm}$. Exclusion criteria were CT scans with inadequate image quality ($n=73$), no typical imaging findings (cavitation, consolidation, centrilobular nodules and tree-in-bud, clusters of nodules, fibronodular scarring, or calcified granulomas) indicative of TB ($n=215$), and negative Mycobacterium tuberculosis culture from sputum, bronchoalveolar lavage, or lung biopsy sample ($n=176$).

CT acquisition parameters and image pre-processing

Chest CT exams were acquired on different CT scanners from multi-centers. The acquisition and reconstruction parameters are summarized in Table 1. CT images were reconstructed with a 512×512 matrix and a slice thickness of $1-1.5\text{ mm}$. Images were preprocessed by setting the lung window (window width, 1500 HU ; window level, -700 HU), and resampled to the voxel to $1 \times 1 \times 1\text{ mm}^3$. A three-dimensional reconstruction approach was used to display the severity of TB visually.

Diagnosis system and network architectures

We proposed a convolutional neural network (CNN)-based AI cascading system to diagnose tuberculosis using a raw chest CT scan. Our model consisted of four subsystems to firstly identify the abnormal CT slices, then to localize the lung region of interest (ROI) and to perform region-specific disease diagnosis, and finally to evaluate the disease severity quantitatively. The cascading networks were automatically connected to create an end-to-end processing pipeline (Fig. 2). All codes are available in the github webpage(<https://github.com/ColinWine/Accurate- and-rapid- pulmonary-tuberculosis-diagnosis-system>). The detailed system network structure for the AI model is summarized in Supplementary Materials.

Image labeling and data annotation

For DL algorithm development (training) and optimization (validation) for lesion localization and classification, a thoracic radiologist (CY) with 11 years of experience performed data annotation. First, the abnormal slices with pulmonary TB lesions and normal slices without pathological findings were manually labeled and used as gold standard to train the DL network. Second, for each CT scan, the TB related imaging features were judged by the Fleischner Society glossary [15], and labeled on the center slices with the maximum area of TB lesions. The bounding boxes of the above critical findings (2.2 lesions per CT scan on average) were drawn by the open-source software ITK-SNAP (version, 3.6.0) for lesion detection and segmentation. Finally, the CT slices and cropped masks (ROIs) from the corresponding lesions area were saved for each scan. For the training phase, up to seven center slices per CT scan were included in patients with multiple lesions.

Slice selection CNN

We designed a binary classification CNN based on Attention Branch ResNet [16] as the slice selection model. Sigmoid function was used as the last activation function (Fig. 3). Two attention mechanisms were introduced to improve its performance. To train the network, 124,890 and 34,161 slices were manually annotated as normal or abnormal, respectively. Slices in each CT scan were ranked according to abnormal scores from the output values of the last Sigmoid function. The top ten abnormal slices from a raw chest CT scan were selected by the algorithm for the subsequent disease diagnosis. We proceed with an evaluation of the ability to screen at the case-level for TB patients vs normal healthy controls. The validation data are provided in Supplementary Materials.

Object detection CNN

The model is based on the CenterNet [6] architecture to provide ROI detection for further classification task [17] (Fig. S1). We used 1,921 bounding boxes with lung pathological features as the input for the model. The ROIs were randomly divided into training and validation cohort in an 8:2 ratio, balancing for imaging findings. The trained CNN can output the location of the TB lesions and the corresponding segmented ROIs. Mean average precision (mAP) was used to evaluate the lesion detection performance of the proposed system.

Disease diagnosis CNN

As for disease diagnosis, we used an 18 layers Squeeze-and-excitation ResNet model (SeNet-ResNet-18) [18] pretrained on the ImageNet dataset to classify detected lesions from selected abnormal CT slices. The network took images with bounding boxes as input and produced a final prediction on categories and activities of TB (Fig. S2). The bounding boxes were labeled as cavitation (n=170), consolidation (n=281), centrilobular nodules and tree-in-bud (n=429), clusters of nodules (n=261), fibronodular scarring (n=264), or calcified granulomas (n=516). We augmented the training data by random rotations range of 0.2 and flipping (horizontal and vertical). A max-pooling layer that outputs maximum probability was used at the last layer.

To further estimate a clinically important task—the activity of TB, we merged the output prediction probability by two-class classification of active (cavitation, consolidation, centrilobular nodules and tree-in-bud, clusters of nodules) and inactive TB (fibronodular scarring, or calcified granulomas) [4]. The label of a patient is then predicted by combining all predictions of every local regions. Max pooling serves as an ‘OR’ gate that labels an image or a case as active TB if there is any subregion that is TB (+).

Severity evaluation

To evaluate the TB burden, we used a Grad-CAM algorithm based on the slice selection CNN described above [16] to generate an activation map. The computerized quantitative approach provided segmentation of the lung tissues based on thresholds and adaptive region growing. Lesion regions were segmented on the “network-activation maps” of all positive slices, which correspond to the region most contributing to the network’s decision. The predefined activation threshold was determined to be 0.88 by visual evaluation by radiologists in our experiments. A lung “TB score” was calculated by the ratio of lesion volumetric summation to that of the corresponding lung lobes.

Additionally, each lobe was visually scored by two independent radiologists with more than 6 years of imaging experience as follows: 0, no lesion; 1, < 5% involvement; 2, 5% to less than 25% involvement; 3, 25% to less than 50% involvement; 3, 50% to less than 75% involvement; and 5, ≥ 75% involvement. Any disagreement between the two radiologists was resolved by another senior radiologist. According to the subjective score of the lung lobes, the patient was defined as severe (greater than or equal to 2) or non-severe (lower than 2).

Performance of the AI model in external validation datasets

Model performance was tested independently at Dataset-2 (Yanling Hospital, n=99), Dataset-3 (Haikou Hospital, n=86) and Dataset-4 [NIH TB Portal dataset (<https://tbportals.niaid.nih.gov/>), n=171]. Patients who matched the above criteria were selected from Dataset-2,3. From the NIH cohort, we then excluded CT scans that fulfilled the following criteria: (a) slice thickness >2mm; (b) inadequate image quality for diagnosis; (c) could not be clearly allocated to either of the classification categories. Model output was probabilities for six critical imaging findings and activity, and the consensus categories by two radiologists with 11 and 6 years of clinical experience were considered as ground truths. Candidate related regions on continuous slices were merged to avoid clustered standard errors. False positive patches were deleted and false negatives were added manually.

Statistical analysis

The confusion matrix was calculated to estimate multiclass classifiers, whereas recall, precision, and the more balanced F1_score were used to measure the performance per-class. The Spearman correlation analysis was performed to assess the correlations between the radiologist-estimated CT score and TB score quantified using the algorithm. The correlation was defined as mild ($r<0.3$), moderate ($0.3\leq r<0.5$), good ($0.5\leq r<0.8$), and strong ($r\geq 0.8$). We also calculated the interobserver reliability for subjective CT score rated from 30 randomly chosen cases by using interclass correlation coefficient. The TB scores of severe and non-severe patients were compared by Student t test. Statistical analysis was performed using SPSS (version 23.0, IBM). Statistical significance was defined as a P value <0.05 .

Results

Patient characteristics

The development dataset included 892 scans from 526 patients (mean age, 48.5 ± 16.5 years; range, 18-92 years; 320 men and 206 women) with clinically diagnosed TB.

Performance of the AI model

Our AI cascading models consisted of four subsystems, which provide consistent visual descriptions: (1) screening to distinguish between normal and abnormal CT images, (2) object detection and localization of pulmonary infectious lesions, (3) diagnostic assessment of radiological features (6 types) and TB activity, (4) severity evaluation.

Slice selection

We performed a slice-level analysis to select the top ten positive slices according to the predicted probability for each CT scan. The average accuracy within the training and validation cohort was 99.6% and 99.8%, respectively.

Object detection and localization

An advanced real-time object detection algorithm based on CenterNet was used to localize lesions, which yielded a mAP of 0.68 on the validation cohort. In the test Dataset-2, 563 candidate regions were detected by the AI model, including 518 true positive (TP), 45 false positive (FP), and 26 false negative (FN). In the test Dataset-3, 502 candidate regions were detected with 440 TP, 62 FP, and 23 FN. In the test Dataset-4, 931 candidate regions were detected with 869 TP, 62 FP, and 37 FN.

Classification and diagnosis

The classification CNN demonstrated a training accuracy of $99.6\%\pm 1.0$ (1531/1537 patches) across six critical imaging findings indicative of TB. For the 20 iterations, the average model accuracy for overall six-way lesion types classification in the validation cohort was 87.37% (346/396). The recall rate for individual lesion types ranged from 83.87% (52/62) for clusters of nodules to 90.00% (90/100) for calcified granulomas. The classification results are summarized as confusion matrix for the critical imaging findings predicted by the AI model (Fig. 4a). In the validation cohort, the DL model showed good discriminative performance for each independent critical finding (area under the curve, AUC = 0.959-0.983, Fig. S3). The classification results of AI model per-class are shown in Table S1.

To validate the general applicability of our AI system, we obtained CT images from an NIH opensource dataset and additional data from our collaborators. In the independent test cohorts (Dataset 2,3,4), the overall classification accuracy for six pulmonary infiltrate types were 91.50% (474/518), 87.65% (440/502) and 86.08% (748/869), respectively. The confusion matrix for each testing dataset is in Fig. 4a. The predictive performance of the corresponding recall, precision and F1-score per-class were listed in Table 2.

For binary decisions, the CNN model achieved an accuracy of 96.97%, 95.35%, and 98.25% for predicting participant-wise activity of the independent test (Dataset-2,3,4) cohorts, respectively (Fig. 4b). The corresponding recall was 100%, 94.87% and 97.87%, precision 95.08%, 94.87% and 98.70%, respectively.

Severity evaluation

A Grad-CAM framework for automatically highlighting pulmonary lesions was used to assess the extent of the disease. The intraclass correlation coefficient for agreement between the two radiologists' subjective score was strong (0.92, 95%CI: 0.90-0.95). As displayed in the attention heatmap obtained by fusion (Fig. 5a), the AI-discovered suspicious infectious area matched highly with the actual pulmonary TB lesions. The Spearman correlation analyses demonstrated a moderate to good correlation between the AI model quantified "TB score" and the radiologist-estimated CT score (r ranged from 0.545 to 0.713) in the validation cohort. The correlation results are summarized in Table 3. The TB scores per lobe in the severe patients were significantly higher than those in the non-severe patients in the validation and testing sets (all $P<0.05$; Fig. 5b). Fig. 6 shows some examples of TB and the corresponding prediction results.

Discussion

In this retrospective, multicohort, diagnostic study, we developed and evaluated an AI cascading model for fully automated diagnosis and tirage of pulmonary tuberculosis based on chest CT images. We found that the model was useful for detection and classification of critical imaging features and achieved an overall accuracy of 0.86–0.92 on external datasets. Moreover, attention heatmap highlighted infectious area for TB burden evaluation with human-level accuracy. The AI system succeeded to stratify patients into severe and non-severe groups by TB score quantified by the algorithm. Our study demonstrated that the DL system allowed for accurate detection, diagnosis, and severity assessment of TB lesions. The proposed AI system can assist clinicians ease the significant demand for pulmonary TB screening, diagnosis and follow-up in daily clinical practice.

TB is the leading cause of death among all infections worldwide, which kills estimated 1.4 million people each year [19]. Despite general incidence has slowly decreasing during the past decade, TB remains an enormous burden globally [3]. The disease is mainly transmitted by cough aerosol, and usually characterized by necrotising granulomatous inflammation in the lung (~ 85% of cases) [20]. Prompt detection, timely treatment, and routine follow-up are priorities to prevent development of TB-related morbidity and spread to the unaffected. Although sputum and blood assays have been developed as the standard to diagnose TB, these tests are inadequate for assessing smear negative TB and results can take up to several days [2]. Lung biopsy provides proven diagnosis pathologically, but is an invasive procedure with significant comorbid risks [21]. Chest radiography is a widely available imaging tools in screening and diagnosis of TB [22]. In comparison, CT is more precise than radiographs in the detection and localization of pulmonary abnormalities [4]. Experimental results have shown that CT scans can aid radiologists review these suspected TB cases when chest radiographs are inconclusive [23]. Furthermore, CT imaging features (including centrilobular nodules, tree-in-bud, consolidation, and cavitation) have demonstrated strong correlation with the positivity and grading of sputum microbiology results [24].

AI has become state of the art for medical image analysis, and plays a role in supporting clinical decision making with respect to diagnosis and risk stratification [7, 25]. Lakhani et al [26] recently reported a deep learning with CNN system to classify TB at chest radiography fast and accurately, with a sensitivity of 97.3% and specificity 100.0%. When evaluating a three-dimensional CT scan, abnormal slices in the full series of images must be identified firstly. To reduce computation burden of AI model, we used a pretrained network to select key slices with the top ten confidence to represent a complete CT scan in this study. We obtained similar accuracy value of 99.80% along with the BConvLSTM U-Net method for image selection in the literature [16]. But this strategy can be with a tradeoff on missing diagnostic information in the slices that are not selected for further analysis.

In lung diseases, several previous studies explored the detection and classification of pulmonary infections, especially during the recent pandemic of COVID-19 [16, 27–29]. Jaeger et al. [30] utilized the a simple but effective one-stage detection model Retina U-Net for lesion detection and localization, achieving a mAP of 0.50. In our work, a CNN based on architecture Centernet was built and trained with human annotations to the inflammatory area. The DL system managed to automatically discover the suspected regions that are strongly indicative of TB with a mAP of 0.68. More recently, Li and colleagues [31] reported a state-of-the-art 3D deep learning model to annotate the spatial location of lesions and classify five critical CT imaging types of TB disease (miliary, infiltrative, caseous, tuberculoma, and cavitary), with classification precision rate at 90.9%. Our model has shown similar overall accuracy (0.86–0.92 vs. 0.91) for six typical CT imaging findings. Another novel development within our AI model is the ability to analyze the imaging features and prediction of activity simultaneously. The increased accuracy for participant-wise prediction (98% compared with region-wise accuracy of 90%) is rational. Errors in region-wise predictions are minimized for the prediction of active TB patients. This could help to more effective identification, intervention and isolation containment of active cases.

Further, a computer-aided CT image analysis tools were recently developed in the diagnosis and evaluation of disease burden for Coronavirus positive patients, which can help to predict the progression to critical illness [14]. Similarly, Shan et al [29] reported a CT-based deep learning system automatically focused on segmentation and quantification of infection regions, which discovered suspicious abnormal areas distributed on bilateral lung in a slice-based “heat map”. This would allow automatic infection delineation and severity prediction consistently and quantitatively. In this study, we demonstrated that the proposed “TB score” corresponds to the disease severity. The lesion percentage determined by radiologists and the AI model showed a moderate to good correlation (r , 0.453–0.761). Furthermore, we demonstrated significant differences in TB score of severe and non-severe groups in the testing datasets (all $P < 0.05$). We therefore hypothesize that such a TB score may provide a quantitative tool for patient follow-up and management to monitor progression and regression of findings. More important, unlike prior supervised algorithms based on slice-level analysis, our approach can search an entire CT study without human guidance. Then the exported quantitative report, with overall TB infection probability, imaging features with spatial coordination, activity and severity prediction may serve as an effective reference for clinicians to make decisions, which is well-suited in real-life health services.

Our study has several limitations. First, the sample size of patients was relatively small, and insufficient data for training the AI networks could limit the performance. CT data from different centers are quite heterogeneous in the scanning parameters and slice thickness. However, such heterogeneous data perhaps allow the results more generalizable. As demonstrated in the independent test cohorts, our AI model is robust to slice thickness and lesion distribution. Second, this study only focused on the six typical types of pulmonary infiltrates but ignored the other rare signs including pleural effusion and enlarged lymph nodes. The sample size of the rare imaging features is not enough for training. Additionally, CT examinations were consecutively acquired, which makes selection bias of different types of imaging findings. Third, Given the overlap of imaging features with other types of pneumonia, our AI systems need to be trained with a larger dataset with abnormalities due to other diseases such as community acquired pneumonia or fungal infections. Lastly, complex deep leaning models are known to face the problem of overfitting. The external validation datasets within our study are inherently different because the cohorts are comprised of patients with different disease burdens. Slight dropout in performance in the test phase was shown in this study.

Conclusions

In conclusion, our deep learning cascading model based on chest CT images can be clinically applicable for accurate detection, diagnosis and triage of TB in the lungs. This fully automated AI system has great potential to be utilized in clinical practice for rapid assessment of disease activity and guidance of therapy and management for patients with pulmonary TB.

Abbreviations

AI: artificial intelligence; CNN: convolutional neural network; COVID-19: Coronavirus disease 2019; CT: computed tomography; AUC: area under the curve; DL: deep learning; FN: false negative; FP: false positive; mAP: mean average precision; ROI: region of interest; TB: tuberculosis; TP: true positive.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethnic Committee at the Nanfang Hospital of Southern Medical University. Informed consent and clinical trial registration were waived, mainly due to the retrospective nature of this study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used or analyzed during the current study are available per reasonable request from the corresponding author.

Competing interests

Dr PL reports, within and outside the submitted work, grants/sponsored research agreements from Varian medical, Oncoradiomics, ptTheragnostic, Health Innovation Ventures and DualTpharma. He received an advisor/presenter fee and/or reimbursement of travel costs/external grant writing fee and/or in-kind manpower contribution from Oncoradiomics, BHV, Merck and Convert pharmaceuticals. Dr PL has shares in the company Oncoradiomics SA and Convert pharmaceuticals SA and is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics and one issue patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patentable invention (software) licensed to ptTheragnostic/ DNAmito, Oncoradiomics and Health Innovation Ventures. Dr HW has (minority) shares in the company Oncoradiomics. LW, TZ and WN serve in the Philips Healthcare. They all provided technical support on deep learning programming. None of the authors had control of the data and information that might present a conflict of interest for the other employee or consultant authors.

Funding

This work was supported by the Natural Science Foundation of Guangdong Province, China (No. 2017A030310102, 2018030310343, and 2020B1515020008) and Medical Scientific Research Foundation of Guangdong Province (No. A2018014). Authors also acknowledge financial support from (ERC-ADG-2015 no 694812-Hypoximmuno), ERC-2020-PoC: 957565-AUTO.DIST ERC advanced grant INCT. Authors also acknowledge financial support from EUROSTARS (COMPACT-12053), the European Union's Horizon 2020 research and innovation programme under grant agreement: ImmunoSABR no 733008, MSCA-ITN-PREDICT no 766276, CHAIMELEON no 952172, EuCanImage no 952103, TRANSCAN Joint Transnational Call 2016 (JTC2016 CLEARLY no UM 2017-8295) and Interreg V-A Euregio Meuse-Rhine (EURADIOMICS no EMR4). This work was supported by the Dutch Cancer Society (KWF Kankerbestrijding), Project number 12085/2018-2.

Authors' contributions

Conceptualization: YX, PL; Data curation: CY, JL, XL, JH; Formal analysis: CY, LW; Funding acquisition: CY, JX, YX, PL; Investigation: JX, JQ, GW, JH; Methodology: JX, JQ, GW; Project administration: TZ, WN, HW; Resources: JL, JX, JH;

Software: LW, WN; Supervision: YX, HW, PL; Validation: TZ, HW; Visualization: GW; Writing—original draft: CY, LW, JL; Writing—review & editing: All authors.

Acknowledgements

Not applicable.

References

1. The global burden of tuberculosis: results from the Global Burden of Disease Study 2015. *Lancet Infect Dis.* 2018; 18(3):261-84.
2. Furin J, Cox H, Pai M. Tuberculosis. *Lancet.* 2019; 393(10181):1642-56.
3. Dheda K, Gumbo T, Gandhi NR, Murray M, Theron G, Udwadia Z, et al. Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis. *Lancet Respir Med.* 2014; 2(4):321-38.
4. Skoura E, Zumla A, Bomanji J. Imaging in tuberculosis. *Int J Infect Dis.* 2015; 32:87-93.
5. Nachiappan AC, Rahbar K, Shi X, Guy ES, Mortani BEJ, Shroff GS, et al. Pulmonary Tuberculosis: Role of Radiology in Diagnosis and Management. *Radiographics.* 2017; 37(1):52-72.
6. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *Eur J Radiol.* 2020; 123:108774.
7. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018; 18(8):500-10.
8. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med.* 2020; 26(6):900-08.
9. Ting D, Cheung CY, Lim G, Tan G, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA.* 2017; 318(22):2211-23.
10. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med.* 2019; 25(6):954-61.
11. England JR, Cheng PM. Artificial Intelligence for Medical Image Analysis: A Guide for Authors and Reviewers. *AJR Am J Roentgenol.* 2019; 212(3):513-19.
12. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. *Radiology.* 2019; 290(3):590-606.
13. Visvikis D, Cheze LRC, Jaouen V, Hatt M. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. *Eur J Nucl Med Mol Imaging.* 2019; 46(13):2630-37.
14. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell.* 2020; 181(6):1423-33.
15. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology.* 2008; 246(3):697-722.
16. Mei X, Lee HC, Diao KY, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med.* 2020; 26(8):1224-28.
17. Li W, Yang Y, Zhang K, Long E, He L, Zhang L, et al. Dense anatomical annotation of slit-lamp images improves the performance of deep learning for the diagnosis of ophthalmic disorders. *Nat Biomed Eng.* 2020; 4(8):767-77.
18. Yasaka K, Akai H, Abe O, Kiryu S. Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study. *Radiology.* 2018; 286(3):887-96.

19. WHO. Global Tuberculosis Report. 2020. <https://apps.who.int/iris/handle/10665/336069>.
20. Dheda K, Barry CR, Maartens G. Tuberculosis. Lancet. 2016; 387(10024):1211-26.
21. Zhou Z, Dong Y, Li C, Wang D, Lv Y, Lv P, et al. Determining the optimal puncture site of CT-guided transthoracic needle aspiration biopsy for the diagnosis of tuberculosis. J Thorac Dis. 2020; 12(8):3987-94.
22. Jaeger S, Karargyris A, Candemir S, Siegelman J, Folio L, Antani S, et al. Automatic screening for tuberculosis in chest radiographs: a survey. Quant Imaging Med Surg. 2013; 3(2):89-99.
23. Kim WS, Choi JI, Cheon JE, Kim IO, Yeon KM, Lee HJ. Pulmonary tuberculosis in infants: radiographic and CT findings. AJR Am J Roentgenol. 2006; 187(4): 1024-33.
24. Ko JM, Park HJ, Kim CH, Song SW. The relation between CT findings and sputum microbiology studies in active pulmonary tuberculosis. Eur J Radiol. 2015; 84(11):2339-44.
25. Ren Z, Hu Y, Xu L. Identifying tuberculous pleural effusion using artificial intelligence machine learning algorithms. Respir Res. 2019;20(1):220.
26. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017; 284(2):574-82.
27. Harmon SA, Sanford TH, Xu S, Turkbey EB, Roth H, Xu Z, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. Nat Commun. 2020; 11(1):4080.
28. Ning W, Lei S, Yang J, Cao Y, Jiang P, Yang Q, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. Nat Biomed Eng. 2020; 4(12):1197-207.
29. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Abnormal Lung Quantification in Chest CT Images of COVID-19 Patients with Deep Learning and its Application to Severity Prediction. Med Phys. 2020. doi:10.1002/mp.14609.
30. Jaeger PF, Kohl SAA, Bickelhaupt S, Isensee F, Kuder TA, Schlemmer HP. Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection. 2018. arXiv:1811.08661.
31. Li X, Zhou Y, Du P, Lang G, Xu M, Wu W. A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis. Appl Intell. 2020. doi:10.1007/s10489-020-02051-1.

Tables

Table 1

Chest CT acquisition parameters of different datasets

| Development dataset | | | Test dataset (Yanling) ^a | | Test dataset (Haikou) |
|----------------------------|-------------------------|----------------------------|--|--------------------|--------------------------|
| CT scanners | Philips Brilliance iCT | Siemens SOMATOM Definition | GE Revolution | Siemens Emotion 16 | Philips Brilliance iCT |
| Case number | 519 | 336 | 37 | 99 | 86 |
| Tube voltage (kVp) | 120 | 110-120 | 100 | 130 | 120 |
| Tube current | automatic mA modulation | automatic mA modulation | automatic mA modulation | 184 mA | automatic mA modulation |
| Pitch | 0.991 | 1.2 | 0.992 | 1.2 | 0.993 |
| Detector configuration(mm) | 128×0.625 | 192×0.6 | 128×0.625 | 192×0.6 | 128×0.625 |
| Resolution | 512×512 | 512×512 | 512×512 | 512×512 | 512×512 |
| Section thickness (mm) | 1 | 1-1.5 | 1.25 | 1.5 | 5 |

^aScanning parameters from NIH open-source dataset are unavailable

Table 2

Performance of the AI model for classification and diagnosis in the test phase

| | Test set (Dataset-2) | | | Test set (Dataset-3) | | | Test set (Dataset-4) | | |
|-------------------------------|----------------------|---------------|----------|----------------------|---------------|----------|----------------------|---------------|----------|
| | Recall (%) | Precision (%) | F1-score | Recall (%) | Precision (%) | F1-score | Recall (%) | Precision (%) | F1-score |
| Cavitation | 88.89% | 100.00% | 94.12% | 86.85% | 100.00% | 92.96% | 86.08% | 95.77% | 90.67% |
| Consolidation | 80.77% | 87.50% | 84.00% | 89.55% | 87.50% | 88.51% | 88.89% | 72.73% | 80.00% |
| Centrilobular and Tree-in-bud | 94.95% | 91.59% | 93.24% | 87.96% | 91.59% | 89.74% | 83.76% | 92.65% | 87.98% |
| Clusters of nodules | 82.14% | 69.70% | 75.41% | 86.67% | 69.70% | 77.26% | 81.32% | 67.27% | 73.63% |
| Fibronodular scarring | 84.44% | 97.44% | 90.48% | 86.35% | 97.44% | 91.56% | 89.55% | 70.59% | 78.95% |
| Calcified granulomas | 92.19% | 94.15% | 93.16% | 88.07% | 94.15% | 91.01% | 88.14% | 94.49% | 91.20% |
| Active/Inactive | 100.00% | 92.68% | 96.20% | 94.87% | 94.87% | 94.87% | 97.87% | 98.98% | 98.42% |

Table 3

Correlation coefficient (*r*) of AI quantified TB score and radiologist-estimated subjective score in lung lobes

| | Validation set | Test set (Dataset-2) | Test set (Dataset-3) | Test set (Dataset-4) |
|-----|----------------|----------------------|----------------------|----------------------|
| LUL | 0.713 | 0.655 | 0.652 | 0.762 |
| LLL | 0.649 | 0.582 | 0.579 | 0.692 |
| RUL | 0.723 | 0.542 | 0.660 | 0.564 |
| RML | 0.624 | 0.531 | 0.600 | 0.402 |
| RLL | 0.545 | 0.534 | 0.580 | 0.503 |

LUL left upper lobe, *LLL* left lower lobe, *RUL* right upper lobe, *RML* right middle lobe, *RLL* right lower lobe

Figures

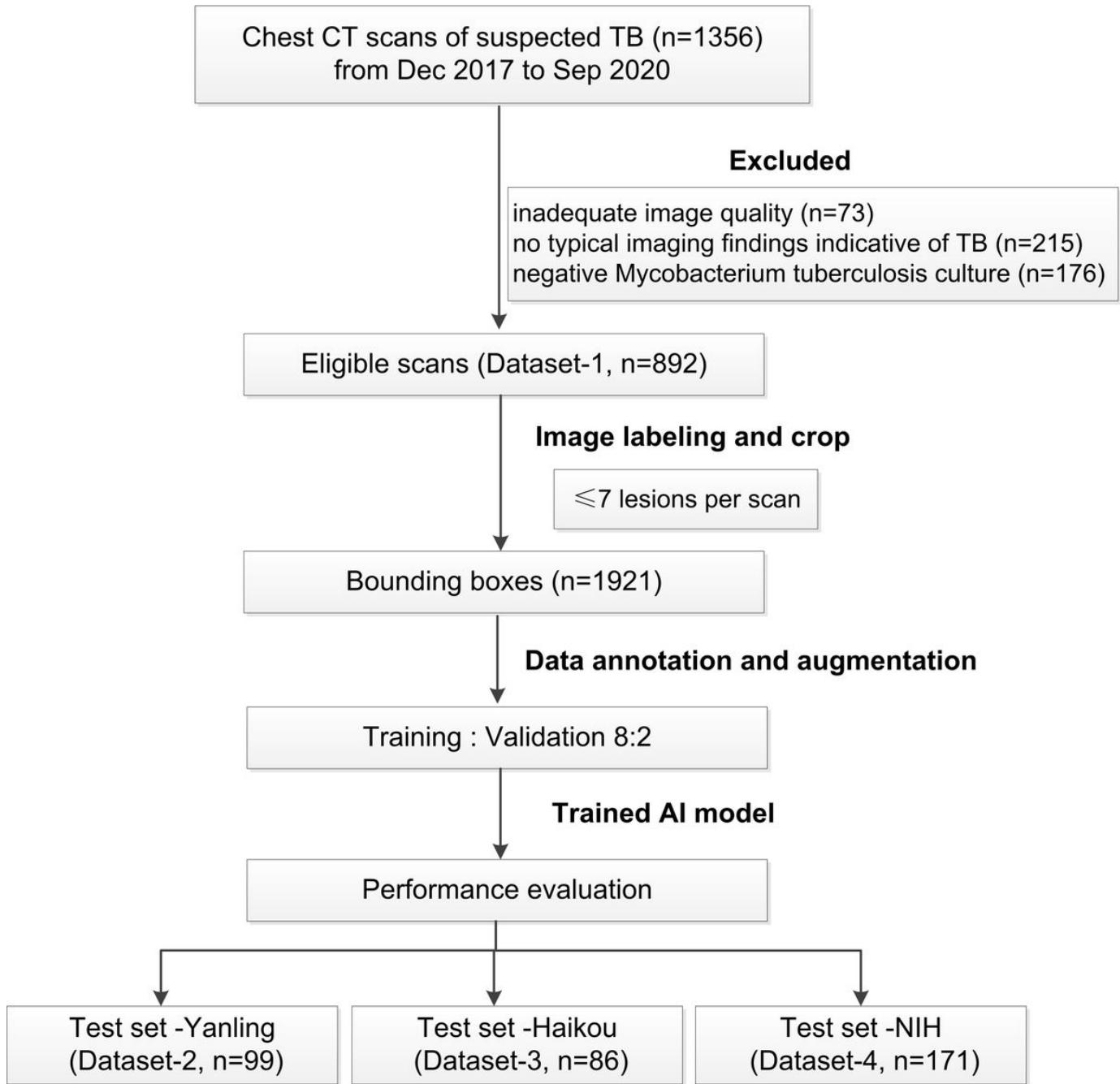


Figure 1

Flowchart of the study process for the training and testing datasets.

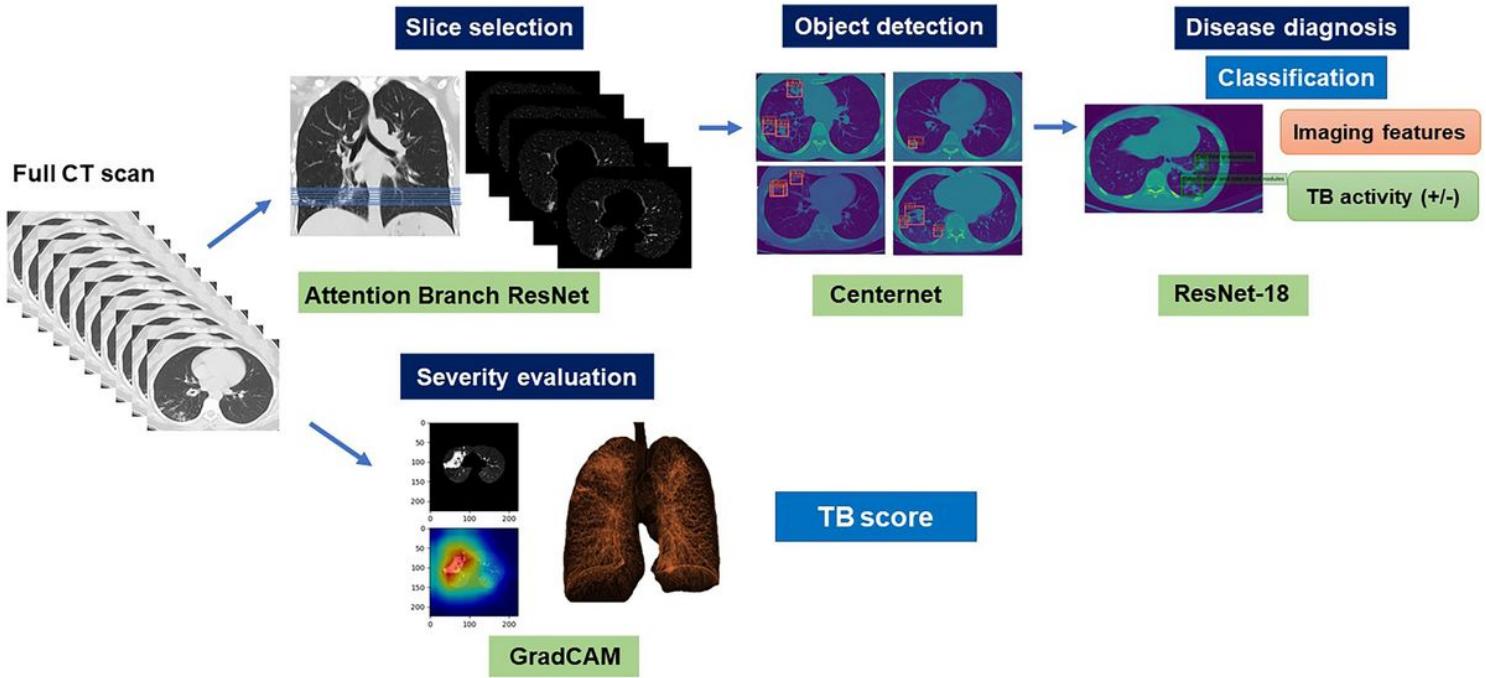


Figure 2

Illustration of the proposed cascading AI pipeline. Our AI diagnostic system consisted of four subsystems, which provides consistent visual descriptions: (1) screening to distinguish between normal and abnormal CT images, (2) object detection and localization of pulmonary infectious lesions, (3) diagnostic assessment of radiological features (6 types) and TB activity, (4) severity assessment. AI artificial intelligence, CT computed tomography, TB tuberculosis.

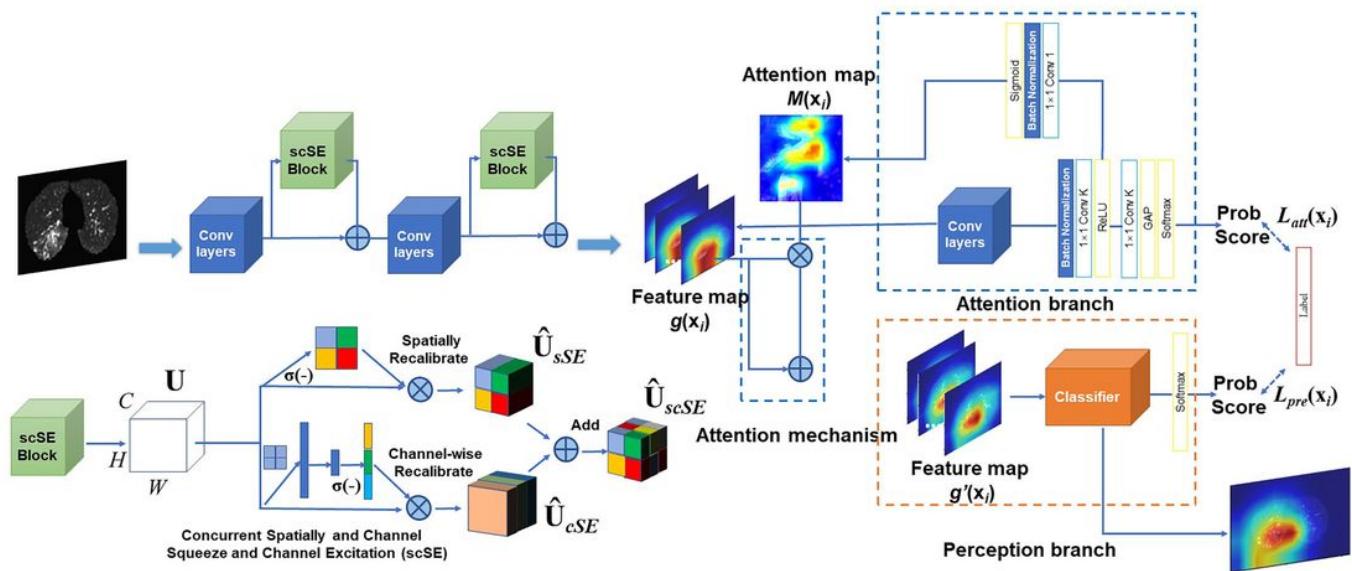


Figure 3

CNN architecture of a deep learning system for slice selection and disease evaluation. Neural network architecture of the subsystem is based on Attention Branch ResNet and GradCAM. The convolution (Conv) layers are used to filter the input full CT scan and extract effective features. CNN convolutional neural network, CT computed tomography.

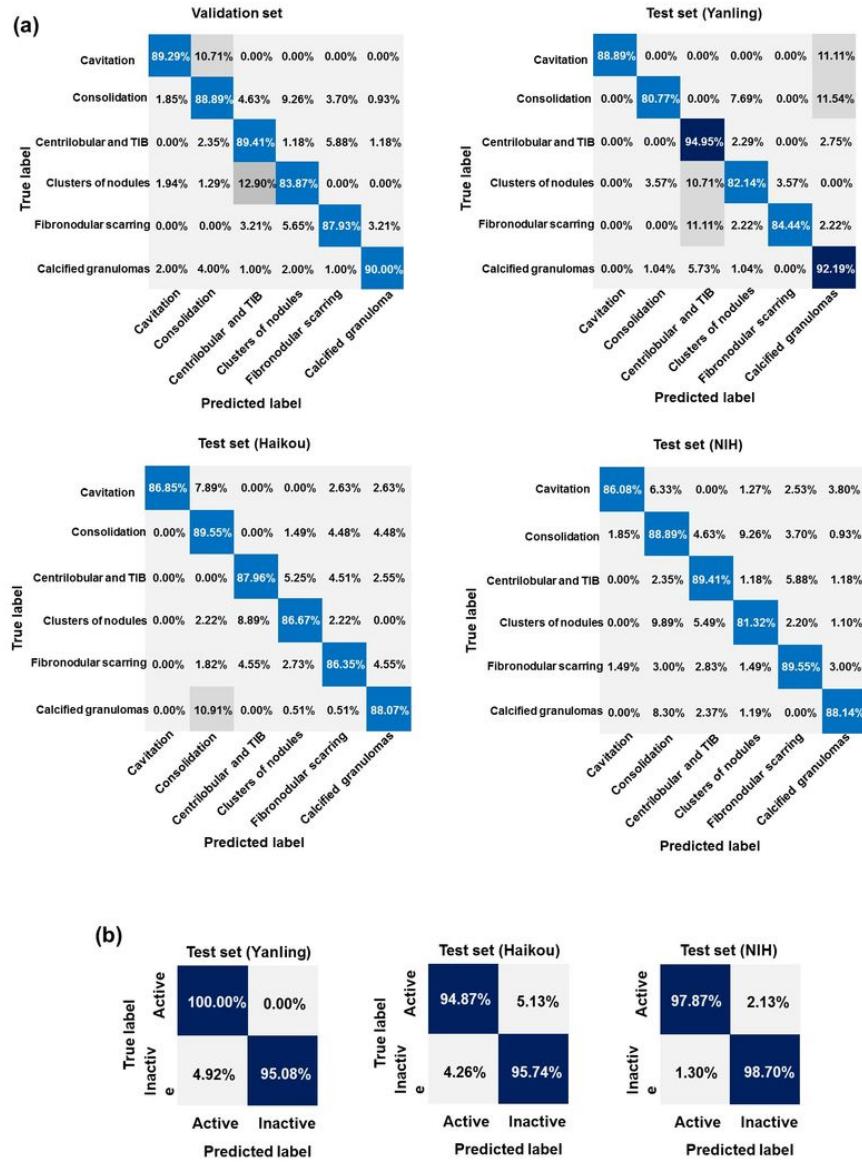


Figure 4

Performance of Our AI System for classification. (a) Normalized confusion matrices of multiclass critical imaging feature classification for the validation and test datasets. (b) Normalized confusion matrices of disease activity prediction for the test datasets. The shaded cells indicate the correct results obtained by the AI system. AI artificial intelligence

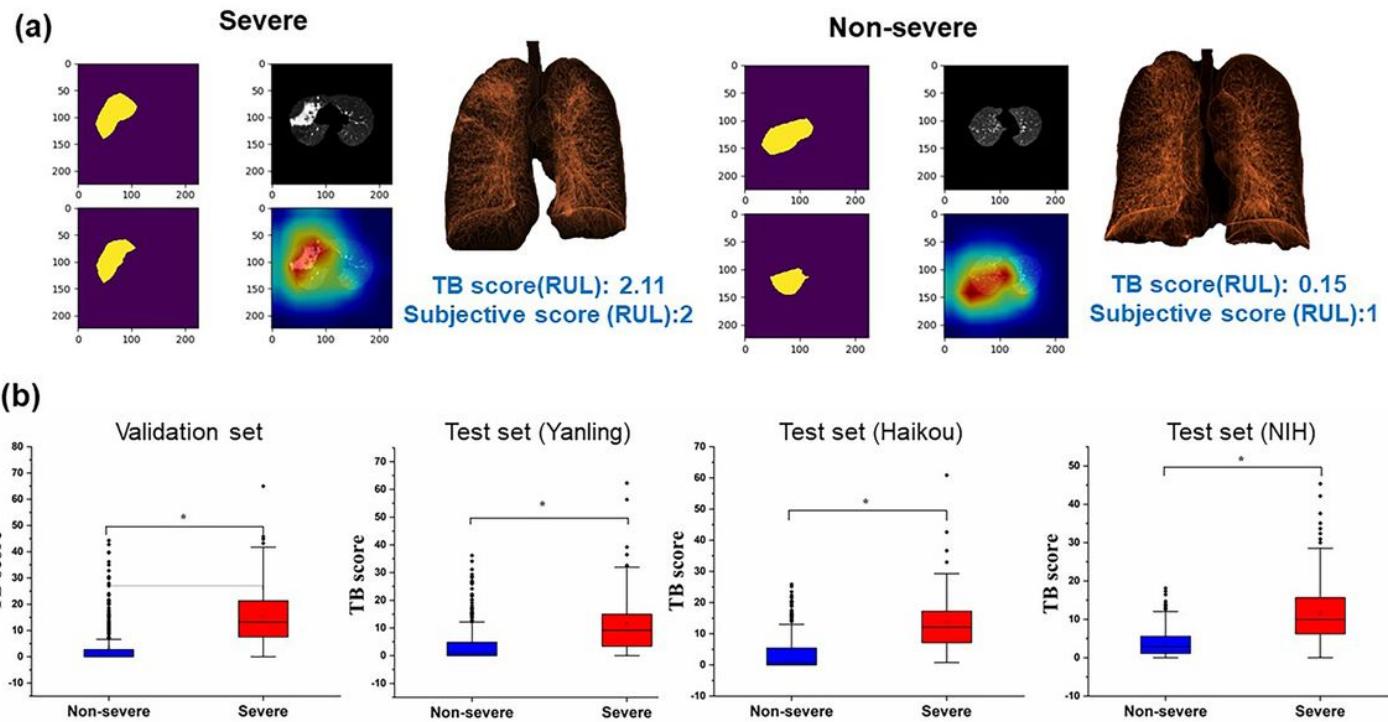


Figure 5

(a) AI discovered suspicious infectious area of severe and non-severe patient. Pseudocolor map represents the three-dimensional reconstruction of the lesion. (b) Boxplots comparing TB scores per-lobe between severe and non-severe patients for the validation and test datasets. AI artificial intelligence, TB tuberculosis

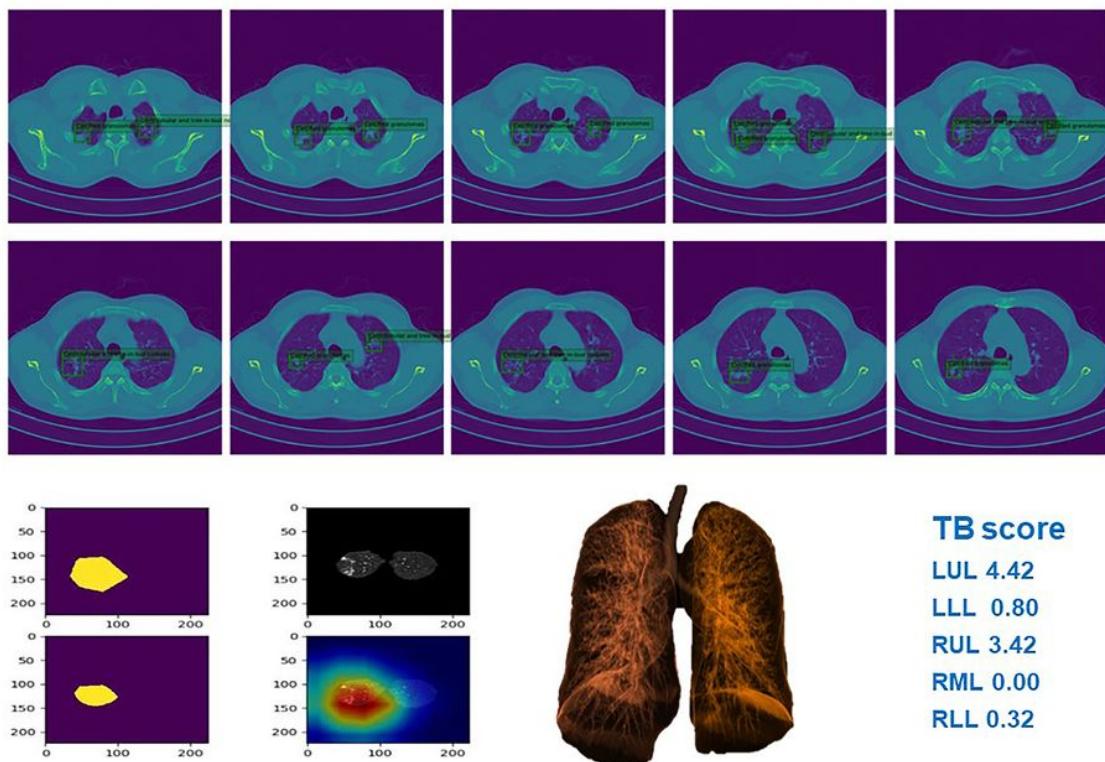


Figure 6

Example of chest CT images of patients with pulmonary TB and performance of our AI model. AI artificial intelligence, CT computed tomography, TB tuberculosis

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.docx](#)