

# A fully automatic artificial intelligence-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis

Citation for published version (APA):

Yan, C. G., Wang, L. F., Lin, J., Xu, J., Zhang, T. J., Qi, J., Li, X. Y., Ni, W., Wu, G. Y., Huang, J. B., Xu, Y. K., Woodruff, H. C., & Lambin, P. (2022). A fully automatic artificial intelligence-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. *European Radiology*, 32(4), 2188-2199. <https://doi.org/10.1007/s00330-021-08365-z>

## Document status and date:

Published: 01/04/2022

## DOI:

[10.1007/s00330-021-08365-z](https://doi.org/10.1007/s00330-021-08365-z)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Download date: 10 May. 2024



# A fully automatic artificial intelligence–based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis

Chengong Yan<sup>1,2</sup> · Lingfeng Wang<sup>3,4</sup> · Jie Lin<sup>1,5</sup> · Jun Xu<sup>6</sup> · Tianjing Zhang<sup>4</sup> · Jin Qi<sup>3</sup> · Xiangying Li<sup>7</sup> · Wei Ni<sup>8</sup> · Guangyao Wu<sup>2,9</sup> · Jianbin Huang<sup>1</sup> · Yikai Xu<sup>1</sup> · Henry C. Woodruff<sup>2,10</sup> · Philippe Lambin<sup>2,10</sup>

Received: 20 June 2021 / Revised: 26 August 2021 / Accepted: 27 September 2021 / Published online: 29 November 2021  
© The Author(s) under exclusive licence to European Society of Radiology 2021

## Abstract

**Objectives** An accurate and rapid diagnosis is crucial for the appropriate treatment of pulmonary tuberculosis (TB). This study aims to develop 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. For training and validation of the model, 1921 lesions were manually labeled, classified according to six categories of critical imaging features, and visually scored regarding lesion involvement as the ground truth. A “TB score” was calculated based on a network-activation map to quantitatively assess the disease burden. Independent testing datasets from two additional hospitals (dataset 2,  $n = 99$ ; dataset 3,  $n = 86$ ) and the NIH TB Portals ( $n = 171$ ) were used to externally validate the performance of the AI model.

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

**Conclusions** The proposed end-to-end AI system based on chest CT can achieve human-level diagnostic performance for early detection and optimal clinical management of patients with pulmonary TB.

## Key Points

- Deep learning allows automatic detection, diagnosis, and evaluation of pulmonary tuberculosis.
- Artificial intelligence helps clinicians to assess patients with tuberculosis.
- Pulmonary tuberculosis disease activity and treatment management can be improved.

**Keywords** Pulmonary tuberculosis · Artificial intelligence · Deep learning · Computed tomography · Thorax

## Abbreviations

AI	Artificial intelligence	FN	False negative
AUC	Area under the curve	FP	False positive
CNN	Convolutional neural network	mAP	Mean average precision
COVID-19	Coronavirus disease 2019	ROI	Region of interest
CT	Computed tomography	TB	Tuberculosis
DL	Deep learning	TP	True positive

✉ Yikai Xu  
yikaixu917@gmail.com

✉ Philippe Lambin  
philippe.lambin@maastrichtuniversity.nl

Extended author information available on the last page of the article

## Introduction

Tuberculosis (TB) is an airborne infectious disease caused by the bacillus *Mycobacterium* that rapidly spreads resulting in significant public health concerns [1, 2]. Although morbidity and mortality associated with

TB have continued to gradually decrease globally, the disease burden remains substantial in endemic countries [3]. Chest imaging plays a crucial role in the workup of patients with pulmonary TB [4]. In particular, computed tomography (CT) has been used to help diagnose, monitor imaging changes, and evaluate the severity of pulmonary TB [5]. On the other hand, the radiographic features can be suggestive of the type and activity of TB [6]. Active TB demonstrates radiologic findings that include cavitation, consolidation, centrilobular nodules and tree-in-bud, or clusters of nodules, while inactive TB is characterized by fibronodular scarring and calcified granulomas [5]. Hence, radiographic monitoring can support the decision making of clinicians for timely isolation and appropriate treatment. If the results of the workup are positive for active TB, initial concomitant-drug therapy is required. However, the complexity of chest CT images is challenging. In countries with sub-optimal health and surveillance systems, underdiagnosis and missed reporting of TB are common problems.

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 increasing application of AI, has made great progress in medical image analysis, including skin disease classification [8], diabetic retinopathy detection [9], and lung cancer screening by chest CT [10]. These DL algorithms can autonomously “learn” to predict features from manually classified initial data. Recent promising advances in CT-based DL systems have demonstrated the potential of AI-assisted radiological diagnostic technology [11–13]. For example, Zhang et al employed a ResNet-18-supervised network for the diagnosis and prediction of triage of patients with coronavirus disease 2019 (COVID-19) during the global pandemic based on CT [14]. Therefore, we hypothesized that end-to-end DL networks can be designed and established by automatic and adaptive feature learning to achieve human expert-level performance for diagnosis and follow-up.

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

## Materials and methods

### Study participants

The study protocol was approved by the institutional review board of Nanfang Hospital of Southern Medical University. Written informed consent was waived due to its retrospective

nature. The chest CT scans of patients with suspected TB acquired from December 2017 to September 2020 were retrieved from the picture archiving and communication system and review. A total of 1356 CT scans of 865 patients (maximum of two scans per patient) were collected who met the following inclusion criteria (Fig. 1): (1) age  $\geq 18$  years; (2) CT examinations for known or suspected primary or secondary TB; and (3) CT slice thickness  $\leq 1.5$  mm. Exclusion criteria were CT scans with inadequate image quality ( $n = 73$ ), no typical imaging findings indicative of TB ( $n = 215$ ), and negative *Mycobacterium tuberculosis* culture from sputum, bronchoalveolar lavage, or lung biopsy samples ( $n = 176$ ).

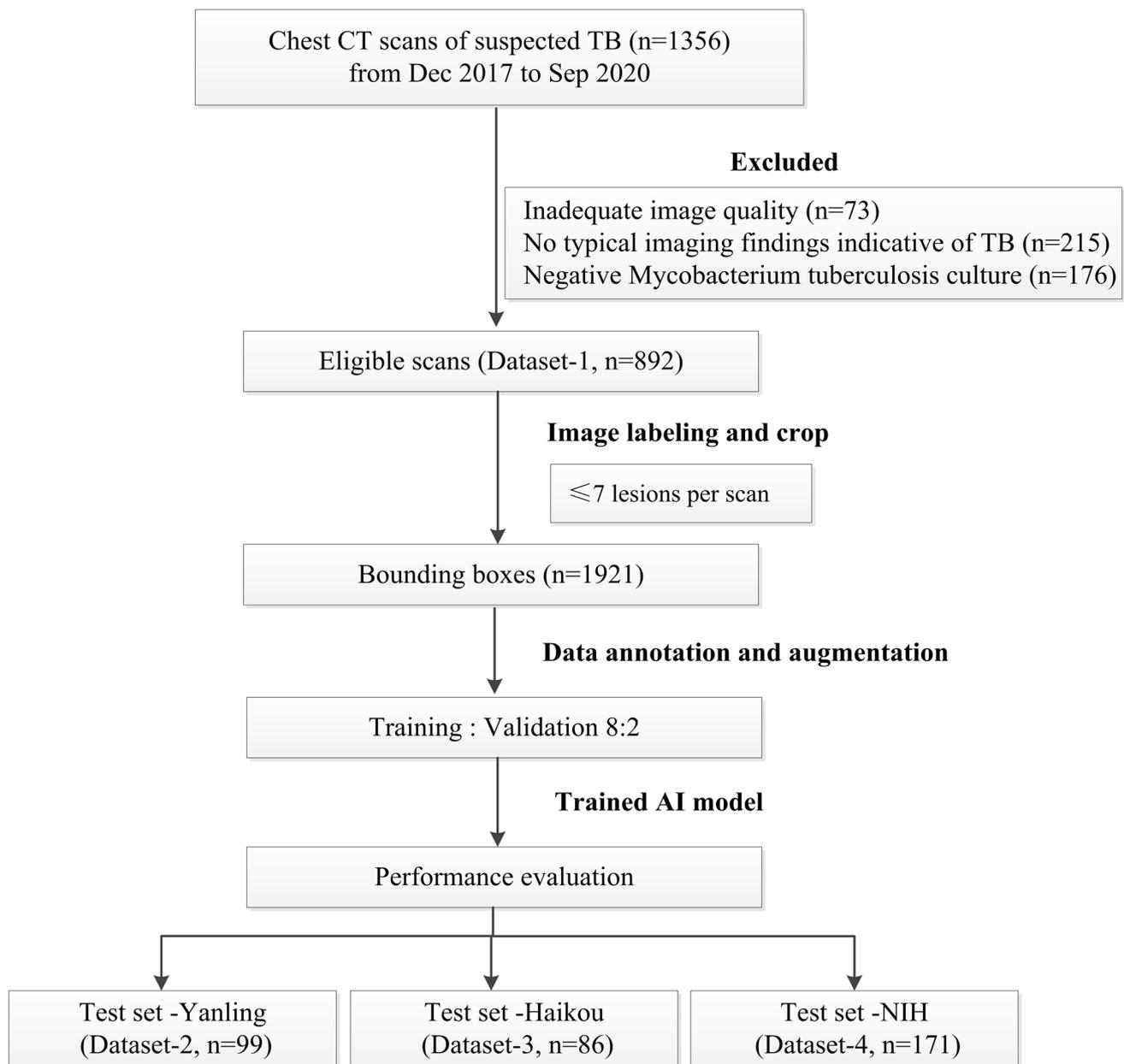
### CT acquisition parameters and image pre-processing

Chest CT was acquired with different CT scanners from multiple centers. The acquisition and reconstruction parameters are summarized in Table 1. CT images were reconstructed with a  $512 \times 512$  matrix and a slice thickness of 1–1.5 mm. Images were preprocessed via the lung window setting (window width, 1500 HU; window level,  $-700$  HU) and resampling the voxels to  $1 \times 1 \times 1$  mm<sup>3</sup>. A three-dimensional reconstruction approach was used to visualize the severity of TB.

### Diagnosis system and network architectures

For DL algorithm development (training) and optimization (validation) for lesion localization and classification, data annotation was performed by a thoracic radiologist (J.H.) with 3 years of experience and verified by an expert radiologist (C.Y.) with 11 years of experience. First, abnormal slices with typical pulmonary TB lesions and normal slices without pathological findings were manually marked and used as gold standards to train the DL network. Second, for each CT scan, the TB-related imaging features were judged according to the Fleischner Society glossary [15], and the center layers with the maximum areas of TB lesions were labeled. The bounding boxes of the above critical findings (average of 2.2 lesions per CT scan) were drawn with the open-source software ITK-SNAP (version, 3.6.0) for lesion detection and segmentation.

Convolutional neural network (CNN)-based cascading networks were automatically connected to create an end-to-end processing pipeline (Fig. 2). The AI model first identified the abnormal CT slices based on the Attention Branch ResNet [16]. The Sigmoid function was used as the last activation function (Fig. 3). The top ten abnormal slices of each raw chest CT scan according to Sigmoid function scores were selected by the algorithm for subsequent analysis. Then, the



**Fig. 1** Flowchart of the study process for the training and testing datasets

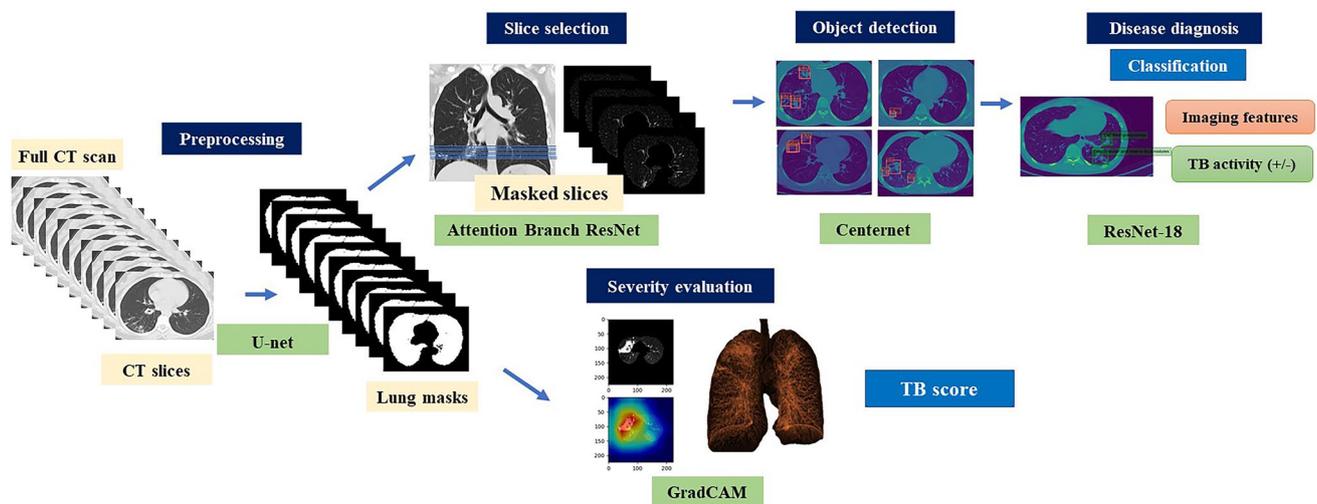
lung region of interest (ROI) based on the CenterNet detection framework was localized [6, 17] (Fig. S1). During the training phase, annotated ground truth ROIs were first converted into heatmaps as input. Then, the trained CNN can output the location of the TB lesions and the corresponding segmented ROIs. CenterNet can generate a heatmap and then the error between the predicted and annotated heatmaps is optimized by the focal loss function. Mean average precision (mAP) was used to evaluate the lesion detection performance of the proposed system. To perform region-specific disease diagnosis, we used an 18-layer squeeze-and-excitation

ResNet model (SeNet-ResNet-18) [18] pretrained on the ImageNet dataset. Images with bounding boxes as input to the network produced a final prediction of the categories and activities of TB (Fig. S2). To further estimate TB activity, the output prediction probability was merged by a two-class classification of active and inactive TB [4]. Finally, a Grad-CAM algorithm based on the slice selection CNN described above [16] was used to generate an activation map to evaluate the disease severity. The computerized quantitative approach provided segmentation of the lung tissues based on thresholds and adaptive region growing. A lung “TB score” was

**Table 1** Chest CT acquisition parameters of different datasets

CT scanners	Development dataset			Test dataset (Yanling)*	Test dataset (Hainan)
	Philips Brilliance iCT	Siemens SOMATOM Definition	GE Revolution		
Scan number	519	336	37	99	86
Tube voltage	120 kVp	110–120 kVp	100 kVp	130 kVp	120 kVp
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	128 × 0.625 mm	192 × 0.6 mm	128 × 0.625 mm	192 × 0.6 mm	128 × 0.625 mm
Resolution	512 × 512	512 × 512	512 × 512	512 × 512	512 × 512
Section thickness	1 mm	1–1.5 mm	1.25 mm	1.5 mm	5 mm

\*Scanning parameters from the NIH open-source dataset are not available



**Fig. 2** Illustration of the proposed cascading AI pipeline. The AI diagnostic system consists of four subsystems, which provides consistent visual descriptions: (1) screening to distinguish between nor-

mal and abnormal CT images, (2) object detection and localization of pulmonary infectious lesions, (3) diagnostic assessment of radiological features (6 types) and TB activity, and (4) severity assessment

calculated by the ratio of lesion volumetric summation to that of the corresponding lung lobes. Additionally, each lobe was visually scored by consensus of two independent radiologists. According to the subjective score, the patient was defined as severe ( $\geq 2$  for any lobe) or non-severe ( $< 2$ ). The detailed system network structure for the AI model is summarized in the [Supplementary Materials](#).

### Performance of the AI model in external validation datasets

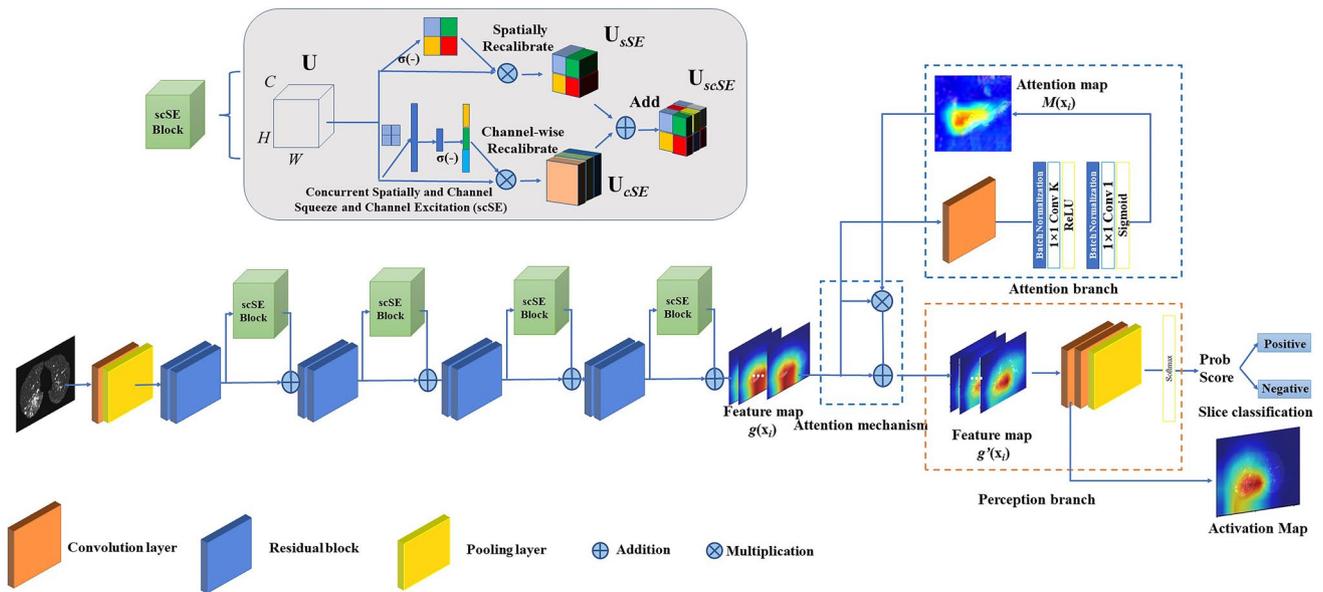
The performance of the model was tested independently at three datasets: dataset 2 (Yanling Hospital,  $n = 99$ ), dataset 3 (Haikou Hospital,  $n = 86$ ), and dataset 4 (National Institutes of Health [NIH] TB Portal dataset at <https://tbportals.niaid.nih.gov/>,  $n = 171$ ). The model output was probabilities for six critical imaging findings and activities, and the consensus categories, as determined by two radiologists (C.Y., J.H.) were considered as ground truths.

mal and abnormal CT images, (2) object detection and localization of pulmonary infectious lesions, (3) diagnostic assessment of radiological features (6 types) and TB activity, and (4) severity assessment

calculated by the ratio of lesion volumetric summation to that of the corresponding lung lobes. Additionally, each lobe was visually scored by consensus of two independent radiologists. According to the subjective score, the patient was defined as severe ( $\geq 2$  for any lobe) or non-severe ( $< 2$ ). The detailed system network structure for the AI model is summarized in the [Supplementary Materials](#).

### Statistical analysis

A confusion matrix was calculated to estimate multiclass classifiers, whereas recall, precision, and a more balanced F1\_score were used to measure the performance by class. Spearman correlation analysis was performed to assess the correlations between the radiologist-estimated CT score and TB score as determined by the algorithm. The correlation was defined as mild ( $r < 0.3$ ), moderate ( $0.3 \leq r < 0.5$ ), good ( $0.5 \leq r < 0.8$ ), or strong ( $r \geq 0.8$ ). Also, the interobserver reliability of the subjective CT score rated from 30 randomly chosen cases was calculated as



**Fig. 3** CNN architecture of a deep learning system for slice selection and disease evaluation. The neural network architecture of the subsystem is based on Attention Branch ResNet and Grad-CAM. The con-

volution (Conv) layers are used to filter the input full CT scan and extract effective features

an interclass correlation coefficient. Student's *t*-test was used to compare the TB scores of severe and non-severe patients. All statistical analyses were performed using IBM SPSS Statistics for Windows, version 23.0. (IBM Corporation). A probability (*p*) value of < 0.05 was considered statistically significant.

## Results

### Patient characteristics

The development dataset included 892 scans of 526 patients (320 men and 206 women; mean age,  $48.5 \pm 16.5$  years; age range, 18–92 years) with clinically diagnosed TB.

### Performance of the AI model

The 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, and (4) evaluation of disease severity.

### Slice selection

Slice-level analysis was performed to select the top 10 positive slices according to the predicted probability for each CT

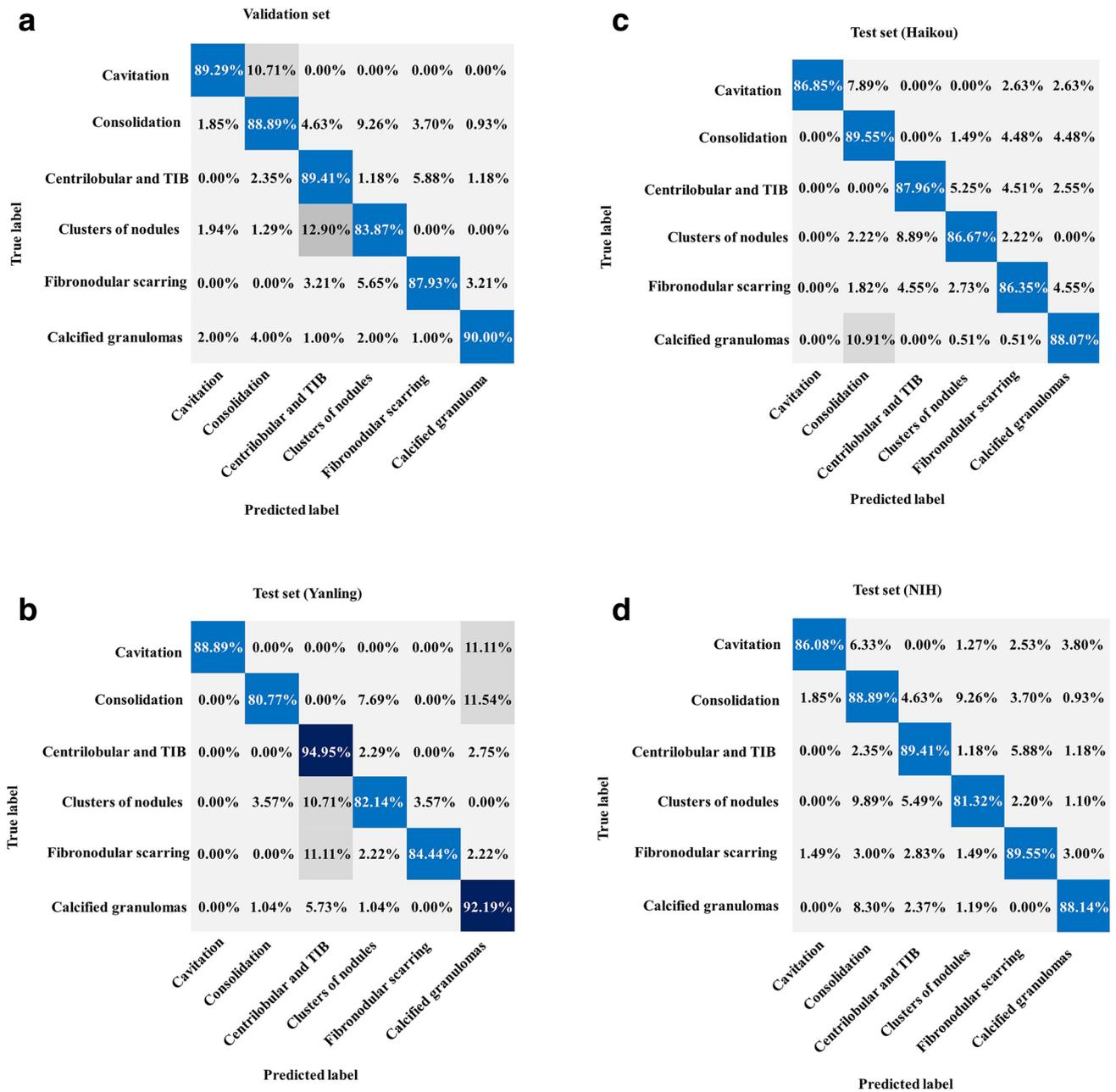
scan. The average accuracy of the training and validation cohort was 99.6% and 99.8%, respectively.

### Object detection and localization

An advanced real-time object detection algorithm based on the CenterNet detection framework was used to localize lesions, which yielded a mAP of 0.68 for the validation cohort. In test dataset 2, 563 candidate regions were detected by the AI model, which included 518 true positives (TPs), 45 false positives (FPs), and 26 false negatives (FNs). In test dataset 3, 502 candidate regions were detected with 440 TPs, 62 FP, and 23 FN. In the test dataset 4, 931 candidate regions were detected with 869 TPs, 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 classification of six-way lesion types in the validation cohort was 87.37% (346/396). The recall rate of individual lesion types ranged from 83.87% (52/62) for clusters of nodules to 90.00% (90/100) for calcified granulomas. The classification results per lesion are summarized as a confusion matrix for the critical imaging findings predicted by the AI model (Fig. 4a). In the validation cohort, the DL model showed good discriminative



**Fig. 4 a–g** Performance of the AI system for classification. **a** Normalized confusion matrices of multiclass critical imaging feature classification for the validation and test datasets. **b** Normalized confu-

sion matrices of disease activity prediction for the test datasets. The shaded cells indicate the correct results obtained by the AI system

performance for each independent critical finding (area under the curve = 0.959–0.983, Fig. S3). The classification results of the AI model by class are shown in Table S1.

To validate the general applicability of the proposed AI system, CT images were obtained from the NIH open-source dataset and additional data from our collaborators. In the independent test cohorts (datasets 2,

3, and 4), the overall classification accuracy rates of the 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 shown in Fig. 4a–d. The predictive performances of the corresponding recall, precision, and F1 score by class are listed in Table 2.

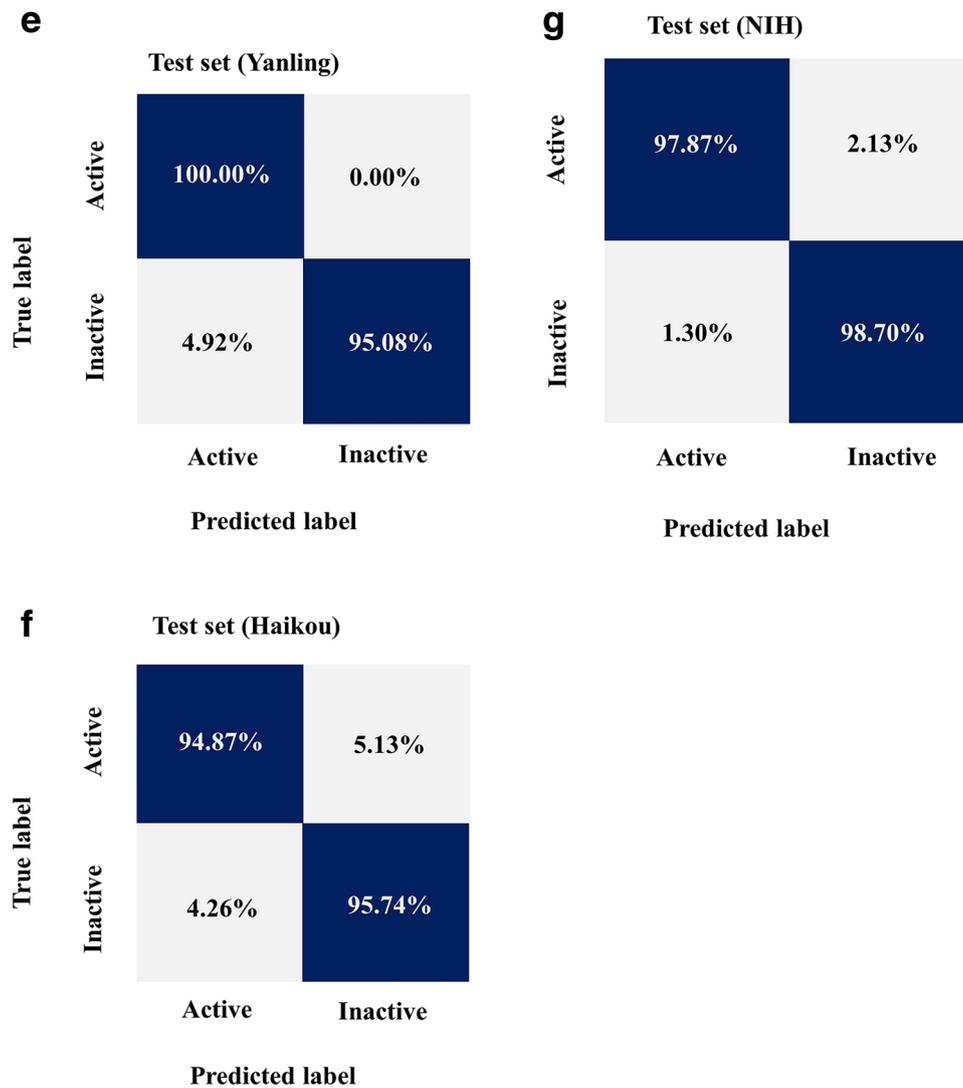


Fig. 4 (continued)

**Table 2** Performance of the AI model for classification and diagnosis in the testing 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%

Note. Data in brackets are 95% confidence interval  
AI artificial intelligence

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 cohorts (dataset 2, 3, and 4), respectively (Fig. 4e–g), while the corresponding recall rate was 100%, 94.87%, and 97.87%, and the precision was 95.08%, 94.87%, and 98.70%, respectively.

### Severity evaluation

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

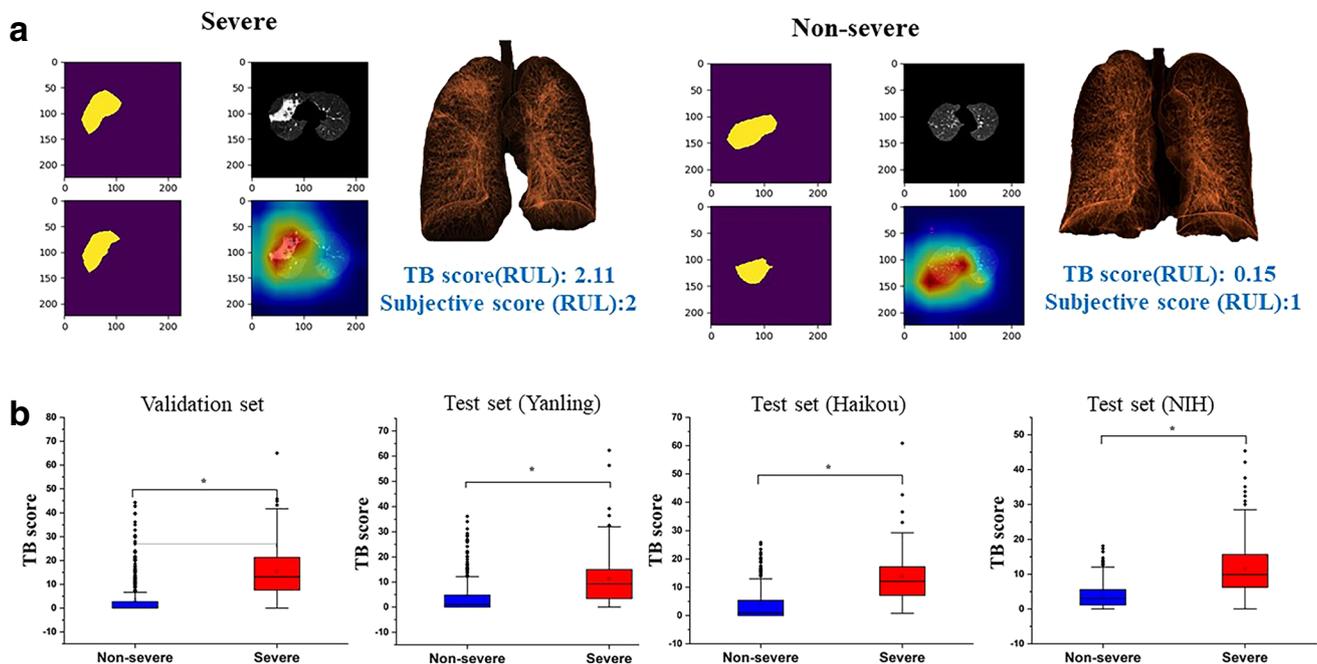
**Table 3** Correlation coefficient ( $r$ ) of AI quantified TB score and radiologist-estimated subjective scores of the lung lobes

	Validation set	Test set (Yanling)	Test set (Haikou)	Test set (NIH)
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

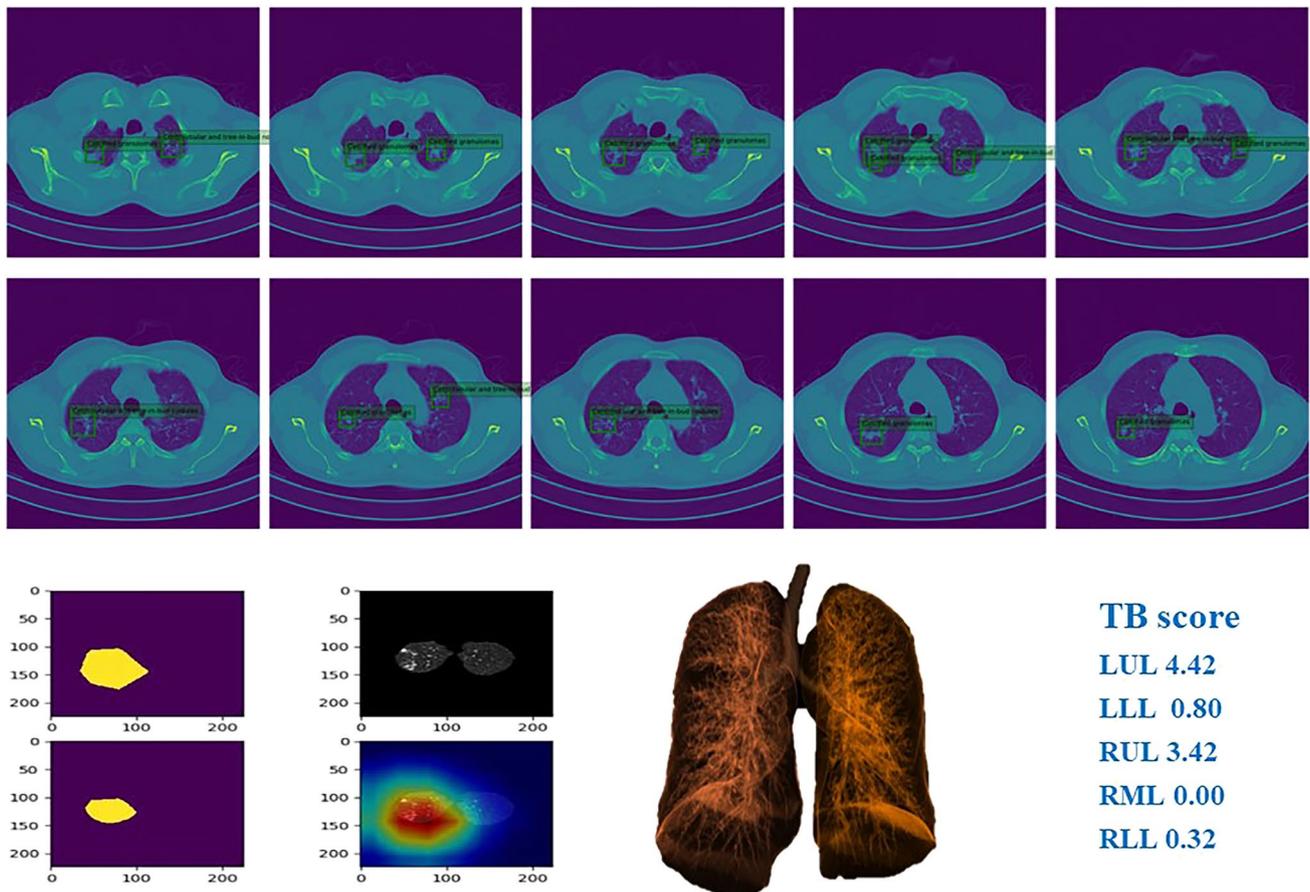
AI artificial intelligence, LUL left upper lobe, LLL left lower lobe, RUL right upper lobe, RML left middle lobe, RLL right lower lobe, TB tuberculosis

### Discussion

In this retrospective, multicohort, diagnostic study, an AI cascading model for fully automated diagnosis and triage of pulmonary TB was developed and evaluated based on chest CT images. The results confirmed that the model was useful for detection and classification of critical imaging features and achieved an overall accuracy of 0.86–0.92 with the use of external datasets. Moreover, an attention heatmap highlighted infectious areas for evaluation of TB burden with human-level accuracy. The AI system succeeded to stratify patients into severe and non-severe groups by TB scores quantified by the algorithm. Furthermore, the DL system allowed for accurate detection, diagnosis, and severity



**Fig. 5** **a** AI-identified suspicious infectious areas on images of severe and non-severe disease. 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



**Fig. 6** Example of chest CT images of patients with pulmonary TB and performance of our AI model

assessment of TB lesions. The proposed AI system can assist clinicians with the significant demands for pulmonary TB screening, diagnosis, and follow-up in daily clinical practice.

TB accounts for an estimated 1.4 million deaths annually worldwide [19]. Although the incidence has continued to gradually decrease over the past decade, TB remains an enormous burden globally [3, 20]. Prompt detection, timely treatment, and routine follow-up are priorities to prevent TB-related morbidity and spread to the unaffected. Although sputum and blood assays are standard for the diagnosis of TB, these tests are inadequate to assess smears negative for TB and the results can take up to several days [2]. Lung biopsy provides a proven diagnosis pathologically, but the procedure is invasive with significant risks for comorbidities [21]. Chest radiography is a widely available imaging tool for screening and diagnosis of TB [22]. Experimental results have shown that CT scans can aid radiologists in the diagnosis of suspected TB cases when chest radiographs are inconclusive [23]. Furthermore, CT imaging features (including centrilobular nodules, tree-in-bud, consolidation, and cavitation) are reported strongly correlated with the positivity and grading of sputum microbiology results [24].

AI has become state of the art for image analysis [25] and plays a role in supporting clinical decision making with respect to diagnosis and risk stratification [7, 26]. Lakhani et al [27] recently reported a DL with CNN system to rapidly and accurately classify TB on chest radiographs with a sensitivity of 97.3% and specificity of 100.0%. When evaluating a three-dimensional CT scan, abnormal slices in the full series of images must be initially identified. To reduce the computation burden of the proposed AI model, a pretrained network was used to select key slices with the top 10 confidence to represent a complete CT scan, which achieved similar accuracy of 99.80% as the BConvLSTM U-Net method for image selection [16]. However, there is a tradeoff with this strategy due to missing diagnostic information from the slices for further analysis.

Several previous studies have explored the detection and classification of pulmonary infections, especially during the recent COVID-19 pandemic [16, 28–31]. Jaeger et al [32] utilized the simple, but effective, one-stage Retina U-Net model for lesion detection and localization, which achieved a mAP of 0.50. In the present study, a CNN based on the Center-Net detection framework managed to automatically identify

suspected regions that were strongly indicative of TB with a mAP of 0.68. More recently, Li and colleagues [33] reported a state-of-the-art three-dimensional DL model to annotate the spatial location of lesions and classify five critical CT imaging types of TB disease (miliary, infiltrative, caseous, tuberculoma, and cavitory), with a classification precision rate at 90.9%. The overall accuracy of the proposed model was similar (0.86–0.92 vs. 0.91) for six typical CT imaging findings. Another novel feature of the AI model is the ability to analyze imaging features and simultaneously predict disease activity. The increased accuracy of 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, which could facilitate more effective identification, intervention, and isolation of active cases.

Further to this, various computer-aided CT image analysis tools were recently developed for the diagnosis and evaluation of the disease burden of coronavirus-positive patients, which can help to predict the progression to critical illness [14]. Similarly, Shan et al [30] reported a CT-based DL system automatically focused on segmentation and quantification of regions of infection, which can identify suspicious abnormal areas of the bilateral lungs as a slice-based “heat map.” This system allows automatic delineation and prediction of disease severity consistently and quantitatively. The results of the present study demonstrated that the proposed TB score corresponds to disease severity. The lesion percentage determined by the AI model was moderate to well correlated to that of the radiologists ( $r = 0.453\text{--}0.761$ ). Furthermore, the AI model demonstrated significant differences in the TB scores between the severe and non-severe groups in the testing datasets (all  $p < 0.05$ ). Therefore, such a TB score provides a potential quantitative tool for patient follow-up and management to monitor progression and regression of findings. More importantly, unlike prior supervised algorithms based on slice-level analysis, the proposed approach can search an entire CT study without human guidance. Then, the exported quantitative report, including overall TB infection probability, imaging features with spatial coordinates, activity, and severity prediction may serve as an effective reference for clinicians to make decisions, which is well-suited in real-life health services.

There were several limitations to this study that should be addressed. First, the sample size was relatively small and the insufficiency of data for training the AI networks may have limited the performance of the model. Second, the CT data from different centers were rather heterogeneous regarding the scanning parameters and slice thicknesses. However, such heterogeneous data may allow the results to be more generalizable. As demonstrated in the independent test cohorts, the proposed AI model is robust regarding slice thickness and lesion distribution. Third, this study only

focused on the six typical types of pulmonary infiltrates, while ignoring other rare signs of primary TB, including pleural effusion and enlarged lymph nodes. The sample size of the rare imaging features was insufficient for training. In the future, the AI system must be trained with a larger dataset with nontypical imaging manifestations to be more suitable for clinical needs. Lastly, overfitting is problematic with complex deep learning models. The external validation datasets used in this study had inherent differences, as the cohorts consisted of patients with different disease burdens, which may explain the slight drop in performance in the testing phase of this study.

In conclusion, the DL cascading model based on chest CT images can be clinically applicable for accurate detection, diagnosis, and triage of pulmonary TB. This fully automated AI system has great potential in clinical practice for rapid assessment of disease activity and guidance of treatment and management of pulmonary TB.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00330-021-08365-z>.

**Funding** This work was supported by the Natural Science Foundation of Guangdong Province, China (nos. 2017A030310102, 2018030310343, and 2020B1515020008) and Medical Scientific Research Foundation of Guangdong Province (no. A2018014); from ERC-ADG-2015 no. 694812-Hypoximmuno and ERC-2020-PoC: 957565-AUTO.DIST ERC advanced grant INCT; and from EUROSTARS (COMPACT-12053), the European Union’s Horizon 2020 research and innovation program under the following 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.

## Declarations

**Guarantor** The scientific guarantor of this publication is Prof. Yikai Xu.

**Conflict of interest** Philippe Lambin 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. PL owns shares in Oncoradiomics SA and Convert Pharmaceuticals SA and is a co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics and one issued patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patentable inventions (software) licensed to ptTheragnostic/ DNAmito, Oncoradiomics, and Health Innovation Ventures. HW owns (minority) shares in Oncoradiomics. Lingfeng Wang, Tianjing Zhang, and Wei Ni have associations with Philips Healthcare, and each provided technical support for DL programming. None of the authors had control of the data that might present a conflict of interest for the other employee or consultant authors.

**Statistics and biometry** One of the authors has significant statistical expertise.

**Informed consent** Written informed consent was waived by the institutional review board.

**Ethical approval** Institutional review board approval was obtained.

#### Methodology

- retrospective
- diagnostic or prognostic study
- multicenter study

## References

1. GBD Tuberculosis Collaborators (2018) The global burden of tuberculosis: results from the global burden of disease study 2015. *Lancet Infect Dis* 18:261–284
2. Furin J, Cox H, Pai M (2019) Tuberculosis. *Lancet* 393:1642–1656
3. Dheda K, Gumbo T, Gandhi NR et al (2014) Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis. *Lancet Respir Med* 2:321–338
4. Skoura E, Zumla A, Bomanji J (2015) Imaging in tuberculosis. *Int J Infect Dis* 32:87–93
5. Nachiappan AC, Rahbar K, Shi X et al (2017) Pulmonary tuberculosis: role of radiology in diagnosis and management. *Radiographics* 37:52–72
6. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP (2020) Artificial intelligence applications for thoracic imaging. *Eur J Radiol* 123:108774
7. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H (2018) Artificial intelligence in radiology. *Nat Rev Cancer* 18:500–510
8. Liu Y, Jain A, Eng C et al (2020) A deep learning system for differential diagnosis of skin diseases. *Nat Med* 26:900–908
9. Ting D, Cheung CY, Lim G et al (2017) Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 318:2211–2223
10. Ardila D, Kiraly AP, Bharadwaj S et al (2019) End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 25:954–961
11. Chassagnon JG, Vakalopolou M, Paragios N, Revel MP (2020) Deep learning: definition and perspectives for thoracic imaging. *Eur Radiol* 30:2021–2030
12. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E (2019) Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology* 290:590–606
13. Visvikis D, Cheze LRC, Jaouen V, Hatt M (2019) Artificial intelligence, machine (deep) learning and radio(gen)omics: definitions and nuclear medicine imaging applications. *Eur J Nucl Med Mol Imaging* 46:2630–2637
14. Zhang K, Liu X, Shen J et al (2020) Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 181:1423–1433
15. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J (2008) Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 246:697–722
16. Mei X, Lee HC, Diao KY et al (2020) Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med* 26:1224–1228
17. Li W, Yang Y, Zhang K et al (2020) Dense anatomical annotation of slit-lamp images improves the performance of deep learning for the diagnosis of ophthalmic disorders. *Nat Biomed Eng* 4:767–777
18. Yasaka K, Akai H, Abe O, Kiryu S (2018) Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology* 286:887–896
19. WHO (2020) Global tuberculosis report. <https://apps.who.int/iris/handle/10665/336069>
20. Dheda K, Barry CR, Maartens G (2016) Tuberculosis. *Lancet* 387:1211–1226
21. Zhou Z, Dong Y, Li C et al (2020) Determining the optimal puncture site of CT-guided transthoracic needle aspiration biopsy for the diagnosis of tuberculosis. *J Thorac Dis* 12:3987–3994
22. Lee JH, Park S, Hwang EJ et al (2021) Deep learning-based automated detection algorithm for active pulmonary tuberculosis on chest radiographs: diagnostic performance in systematic screening of asymptomatic individuals. *Eur Radiol* 31:1069–1080
23. Kim WS, Choi JI, Cheon JE, Kim IO, Yeon KM, Lee HJ (2006) Pulmonary tuberculosis in infants: radiographic and CT findings. *AJR Am J Roentgenol* 187:1024–1033
24. Ko JM, Park HJ, Kim CH, Song SW (2015) The relation between CT findings and sputum microbiology studies in active pulmonary tuberculosis. *Eur J Radiol* 84:2339–2344
25. Agrawal P, Chaudhary D, Madaan V, Zabrovskiy A, Timmerer C (2021) Automated bank cheque verification using image processing and deep learning methods. *Multimedia Tools and Applications* 80:5319–5350
26. Savadjiev P, Chong J, Dohan A et al (2019) Demystification of AI-driven medical image interpretation: past, present and future. *Eur Radiol* 29:1616–1624
27. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284:574–582
28. Harmon SA, Sanford TH, Xu S et al (2020) Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun* 11:4080
29. Ning W, Lei S, Yang J et al (2020) Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. *Nat Biomed Eng* 4:1197–1207
30. Shan F, Gao Y, Wang J et al (2020) Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. *Med Phys* 48:1633–1645
31. Madaan V, Roy A, Gupta C et al (2021) XCOVNet: chest X-ray image classification for COVID-19 early detection using convolutional neural networks. *New Gener Comput* 24:1–15
32. Jaeger PF, Kohl SAA, Bickelhaupt S, Isensee F, Kuder TA, Schlemmer HP (2018) Retina U-Net: embarrassingly simple exploitation of segmentation supervision for medical object detection. *arXiv:1811.08661*
33. Li X, Zhou Y, Du P, Lang G, Xu M, Wu W (2020) A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis. *Appl Intell* 51:4082–4093

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Chenggong Yan<sup>1,2</sup> · Lingfeng Wang<sup>3,4</sup> · Jie Lin<sup>1,5</sup> · Jun Xu<sup>6</sup> · Tianjing Zhang<sup>4</sup> · Jin Qi<sup>3</sup> · Xiangying Li<sup>7</sup> · Wei Ni<sup>8</sup> · Guangyao Wu<sup>2,9</sup> · Jianbin Huang<sup>1</sup> · Yikai Xu<sup>1</sup> · Henry C. Woodruff<sup>2,10</sup> · Philippe Lambin<sup>2,10</sup>

<sup>1</sup> Department of Medical Imaging Center, Nanfang Hospital, Southern Medical University, No. 1838 Guangzhou Avenue North, Guangzhou 510515, Guangdong, China

<sup>2</sup> The D-Lab, Department of Precision Medicine, GROW-School for Oncology and Developmental Biology, Maastricht University, 6229ER, Maastricht, The Netherlands

<sup>3</sup> School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

<sup>4</sup> Clinical and Technical Solution, Philips Healthcare, Guangzhou 510055, China

<sup>5</sup> Department of Radiology, Nanfang Yanling Hospital, Guangzhou 510507, China

<sup>6</sup> Department of Hematology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China

<sup>7</sup> Department of Radiology, Affiliated Haikou Hospital of Xiangya Medical College, Central South University, Haikou 570208, China

<sup>8</sup> Philips China Innovation Hub, Shanghai 200072, China

<sup>9</sup> Present Address: Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

<sup>10</sup> Department of Radiology and Nuclear Imaging, GROW-School for Oncology and Developmental Biology, Maastricht University Medical Centre, 6229ER, Maastricht, The Netherlands