European Journal of Nuclear Medicine and Molecular Imaging (2021) 48:4293–4306
https://doi.org/10.1007/s00259-021-05432-x

ORIGINAL ARTICLE

Distinguishing nontuberculous mycobacteria from *Mycobacterium tuberculosis* lung disease from CT images using a deep learning framework

Li Wang[1] · Wenlong Ding[1] · Yan Mo[2] · Dejun Shi[2] · Shuo Zhang[1] · Lingshan Zhong[1] · Kai Wang[1] · Jigang Wang[1] · Chencui Huang[2] · Shu Zhang[2] · Zhaoxiang Ye[3] · Jun Shen[1] · Zhiheng Xing[1]

Received: 10 February 2021 / Accepted: 24 May 2021 / Published online: 16 June 2021

Abstract

Purpose To develop and evaluate the effectiveness of a deep learning framework (3D-ResNet) based on CT images to distinguish nontuberculous mycobacterium lung disease (NTM-LD) from *Mycobacterium tuberculosis* lung disease (MTB-LD).
Method Chest CT images of 301 with NTM-LD and 804 with MTB-LD confirmed by pathogenic microbiological examination were retrospectively collected. The differences between the clinical manifestations of the two diseases were analysed. 3D-ResNet was developed to randomly extract data in an 8:1:1 ratio for training, validating, and testing. We also collected external test data (40 with NTM-LD and 40 with MTB-LD) for external validation of the model. The activated region of interest was evaluated using a class activation map. The model was compared with three radiologists in the test set.
Result Patients with NTM-LD were older than those with MTB-LD, patients with MTB-LD had more cough, and those with NTM-LD had more dyspnoea, and the results were statistically significant ($p < 0.05$). The AUCs of our model on training, validating, and testing datasets were 0.90, 0.88, and 0.86, respectively, while the AUC on the external test set was 0.78. Additionally, the performance of the model was higher than that of the radiologist, and without manual labelling, the model automatically identified lung areas with abnormalities on CT > 1000 times more effectively than the radiologists.
Conclusion This study shows the efficacy of 3D-ResNet as a rapid auxiliary diagnostic tool for NTB-LD and MTB-LD. Its use can help provide timely and accurate treatment strategies to patients with these diseases.

Keywords Nontuberculous mycobacterium · *Mycobacterium tuberculosis* · Deep learning · Computed tomography · Man–machine comparison

Introduction

Mycobacterium tuberculosis (MTB) is the most common cause of mortality among infections of the respiratory tract, with a reported 1.4 million deaths in 2019 [1]. However, much attention has been focused on *Mycobacterium tuberculosis* lung disease (MTB-LD), which has led to

Li Wang, Wenlong Ding, and Yan Mo contributed equally to this work.

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence).

✉ Jun Shen
hhyy_shenjun@outlook.com

✉ Zhiheng Xing
18920696025@189.cn

[1] TCM Key Research Laboratory for Infectious Disease Prevention for State Administration of Traditional Chinese Medicine, Tianjin Institute of Respiratory Diseases, Haihe Hospital, Tianjin University, Tianjin, People's Republic of China

[2] Deepwise AI Lab, Beijing Deepwise& League of PHD Technology Co., Ltd, Haidian District, 21st Floor, China Sinosteel Plaza, NO. 8, Haidian Avenue, Beijing 100080, People's Republic of China

[3] National Clinical Research Centre of Cancer, Key Laboratory of Cancer Prevention and Therapy, Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, People's Republic of China

underestimation of the prevalence of nontuberculous mycobacterium lung disease (NTM-LD) [2]. According to epidemiologic data, an increase in the prevalence of NTM-LD has been observed in the USA [3] and South Korea [4]; however, no large-sample NTM-LD epidemiologic studies have been conducted in China [5, 6]. The isolation rate of NTM in different parts of China is generally increasing [7–9], and this indicates that the prevalence of NTM-LD in China is on the rise [10]. The main route of transmission of MTB is through inhalation of aerosols from patients with active MTB-LD. When managing patients with MTB-LD, it is necessary to consider the control of aerosol-borne infection. On the other hand, NTM-LD is an opportunistic pathogen that is commonly found in the environment [11]; however, it rarely causes human diseases unless the host is immunocompromised [12]. Therefore, control of aerosol generation by patients with NTM-LD is unnecessary [13]. Considering these issues, accurate and timely diagnosis of the two diseases can contribute to the management of patients.

The differential diagnosis of MTB-LD and NTM-LD poses a challenge to clinicians. The purified protein derivative test is the most common screening method for patients with suspected MTB; however, this test cannot distinguish between MTB-LD and NTM-LD [14]. Furthermore, acid-fast bacilli (AFB) smears of both diseases show positive results [15], while the use of bacterial cultures and strain identification is time-consuming. In many tuberculosis-endemic areas, patients who have a positive AFB smear receive empirical anti-tuberculosis treatment while waiting for the result of bacterial culture. However, patients with NTM-LD will not benefit from this management and can have drug-related adverse effects and incur unnecessary medical expenses. The unresponsiveness of patients with NTM-LD to anti-tuberculosis medications indicates a difference in drug sensitivity between NTM-LD and MTB-LD [16]. Delayed diagnosis of NTM-LD leads to a longer treatment time, prolonged disease course, poor prognosis, and treatment failure. Therefore, accurate diagnosis determines correct treatment, which could help avoid unnecessary drug-related adverse reactions and medical expenses.

Chest radiography is a common screening method for inchoate tuberculosis; however, it cannot be used to distinguish between NTM-LD and MTB-LD [17]. Compared to chest radiography, chest computed tomography (CT) scan can provide more information and clearly show changes in lung areas including cavity formation, parenchymal lesion, bronchiectasis, and tree-in-bud patterns[18] (Fig. 1). However, these changes can be seen in both NTM-LD and MTB-LD; therefore, CT scan alone is insufficient in differentiating between both diseases [19]. Genetic studies [20, 21] have indicated that similarities in the manifestations and imaging results of the two diseases are caused by their similar gene strains. Bronchiectasis in the right middle lobe and left upper lobe lingual segment is considered an important feature of NTM-LD [22, 23]; however, this feature is not rare in patients with MTB-LD [24]. On the other hand, pleural effusion is rare in the CT images of patients with NTM-LD [25]. A 2014 study found that 13.3% of patients with NTM-LD had pleural effusion [26], which may lead to the misdiagnosis of NTM-LD. The difficulty in the distinguishing between the CT images of the two diseases is caused by factors including the experience of the physician, time, and effort [19, 22, 27].

Although chest CT findings are not reliable for distinguishing between NTM-LD and MTB-LD [19], CT aids in diagnosing NTM-LD and MTB-LD before definite mycobacterial culture because of its great availability and short examination time [28–30]. CT images could show the extent and severity of lesion involvement, and evaluate the efficacy of medication through re-examination.

In recent years, artificial intelligence (AI) technology represented by machine learning has increasingly gained popularity in medical imaging. The use of AI technology is expected to meet the challenge posed by the discrimination of NTM-LD and MTB-LD by improving diagnostic accuracy, and sensitivity, leading to a reduction in the time taken to make the correct diagnosis. Three dimensions-residential network (3D-ResNet) is a classic deep learning framework that is highly recognised for its performance and efficiency. Previous studies have shown that 3D-ResNet achieved surprising results in various medical scenarios, such as recessive ischemic stroke recognition [25], Alzheimer's disease classification [31], thoracic vertebral segmentation [32], and colonoscopic polyp detection [33], with an accuracy range of 0.83 from 0.97 in the test set. Additionally, 3D-ResNet has contributed significantly in the diagnosis of lung diseases, such as corona virus disease 2019 (COVID-19) diagnosis using CT scan or chest X-ray (CXR) [34, 35], malignancy risk classification of lung nodules using CT scan [36], non-small cell lung cancer [37], lung nodules detection [38], and lung CT image classification [39], with an accuracy ranging from 0.80 to 0.99 on the test set. Moreover, to the best of our knowledge, this is the first attempt that makes use of deep learning based on 3D-ResNet to distinguish between NTM-LD and MTB-LD. At the same time, we also supplemented a man–machine experiment to evaluate the clinical applicability of the deep learning framework.

In our study, a deep learning framework of NTM-LD and MTB-LD based on chest CT images was developed to assist radiologists in diagnosing these diseases early in order to allocate appropriate treatment strategies to patients with these diseases.

Materials and methods

Patients and dataset

This study retrospectively collected data from patients with pulmonary infection that were admitted to Tianjin Haihe Hospital (Tianjin University, China) between January 2014 and January

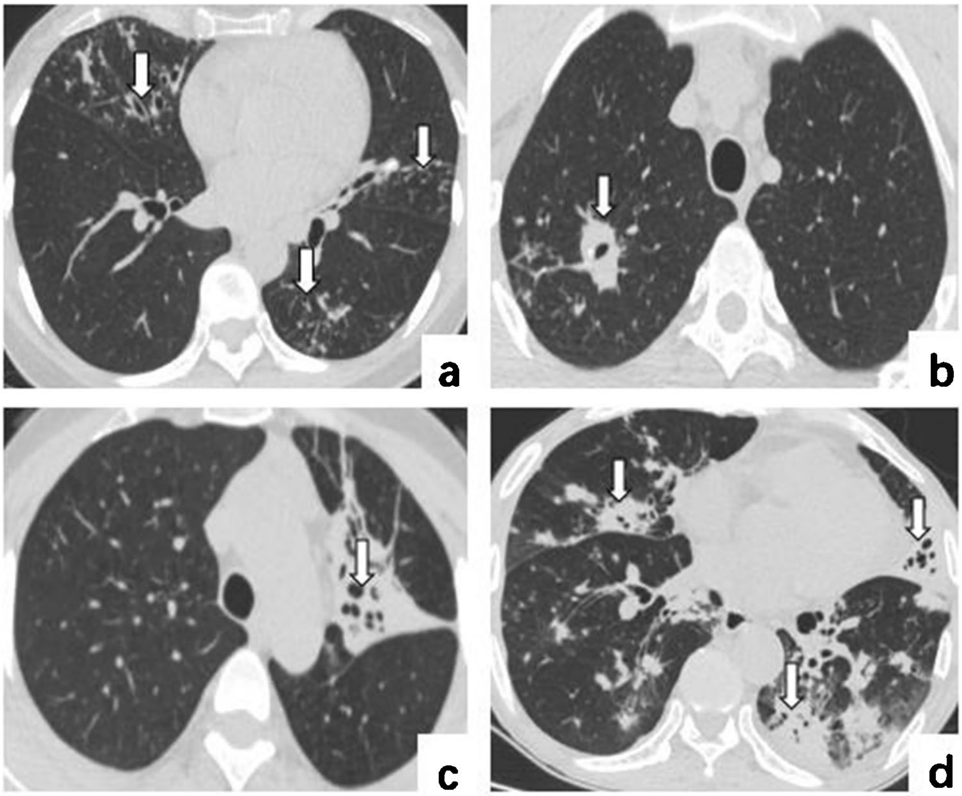


Fig. 1 **a** Correct diagnosis NTM-LD: female, 50 years old, hemoptysis for 3 days, bronchiectasis in right middle lobe, left upper lobe lingual segment and left lower lobe (arrows) with multiple central lobular nodules and tree-in-bud around. **b** Correct diagnosis MTB-LD: male, 24 years old, cough and expectoration for 6 days, thick-walled cavity (arrow) in right upper lobe with smooth inner wall and local calcification of cavity wall, adjacent pulmonary nodules and tree-in-bud around. **c** NTM-LD misdiagnosed as MTB-LD: female, 30 years old, cough and expectoration for more than one year, irregular consolidation in left upper lobe with bronchiectasis (arrow). **d** MTB-LD misdiagnosed as NTM-LD: male, 58 years old, cough, expectoration and fever for 2 weeks, multiple consolidation and bronchiectasis in right middle lobe, left upper lobe lingual segment and lower lobe of both lungs (arrows), and multiple nodules in right lung.

2020, who underwent non-contrast CT examination and were infected by nontuberculous mycobacterium (NTM) or MTB.

Pathogenic microbiological examination was used as the reference in the diagnosis of NTM or MTB infection. Sputum samples were collected, and an AFB stain was performed. Mycobacterium culture was grown using Löwenstein-Jensen medium. NTM species were identified using matrix-assisted laser desorption ionisation-time of flight (MALDI-TOF) mass spectrometry. Specifically, a diagnosis of NTM-LD was based on the treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline (2020) [40], and a diagnosis of MTB-LD patients was based on the National Health Commission of the People's Republic of China. In: diagnostic criteria for pulmonary tuberculosis (WS 288–2017) [41].

We excluded patients who were diagnosed with both diseases, patients with other lung diseases (including infectious diseases, tumours, and interstitial lesions), patients with a history of lung surgery, or patients with poor CT image quality caused by respiratory motion and metal artefacts.

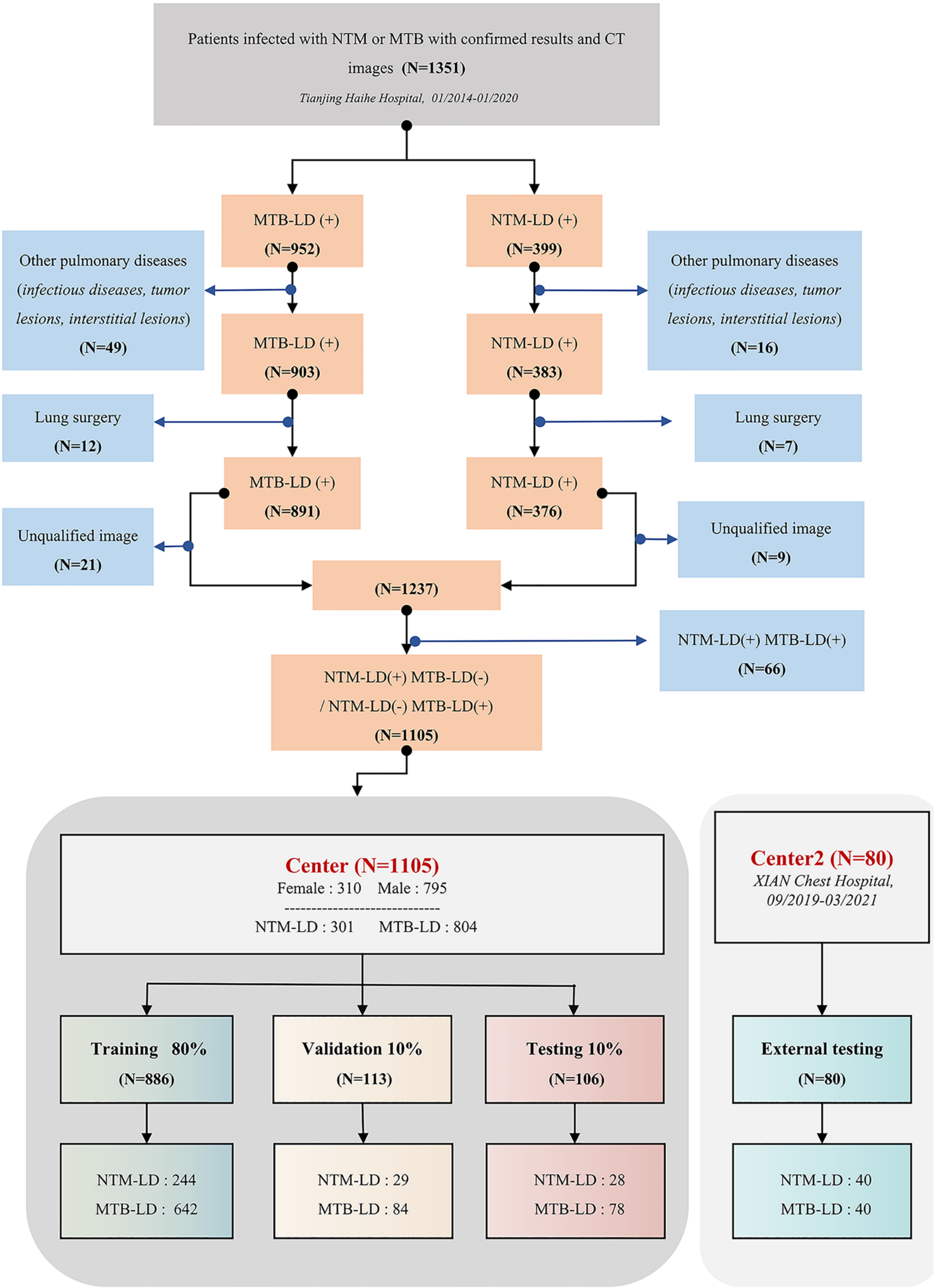
A complete flow chart of the data collection is shown in Fig. 2. A total of 1105 patients were included in this study, comprising 301 patients with NTM-LD and 804 patients with MTB-LD. The patients were randomly divided into three datasets for training, validating, and testing of a deep learning model (3D-ResNet) at a ratio of 8:1:1. Clinical information was collected from each patient, including age, sex, and the presence of symptoms, such as chest pain, cough, expectoration, fever, chest tightness, haemoptysis, gasp, dyspnoea, chills, fatigue, night sweats, and weight loss. CT images and clinical manifestations were collected within 1 month before samples were collected for pathogenic microbiological examination.

We also supplemented the experimental results of the external test set. Following the data inclusion criteria of this study, we retrospectively collected 40 NTM-LD patients and 40 MTB-LD patients that were admitted to Xi'an Chest Hospital (Shanxi, China) between September 2019 and March 2021 to form the external test set.

Image data

The patients in this study were scanned using two spiral CT scanners (BrightSpeed, GE, America and Aquilion Prime 128, Canon Medical Systems, Otawara, Japan) under the same protocol. Prior to scanning, patients assumed a supine position and were instructed to inhale maximally and hold their breath to ensure the accuracy of the data during breathing phase. Afterwards, both lungs were scanned from the apices to the bases. FOV was adjusted according to the patient's body shape. The CT scanning parameters of BrightSpeed with in-plane size of 512 × 512 were as follows: 120 kV, automatic tube current, 0.75 s/r, collimator width = 16 × 0.625 mm, which were reconstructed based on the standard and lung algorithm (thickness of 1.25 mm, interval of 1.25 mm). The voxel size of BrightSpeed was 0.31–1.14 mm^3. The CT scanning parameters of Aquilion Prime 128 with an in-plane size of 512 × 512 were as follows: 120 kV, automatic tube current, 0.5 s/r, collimator width = 64 × 0.5 mm, which were reconstructed based on the FC 30 and FC 52 algorithm (thickness of 1.0 mm, interval of 0.8 mm). The voxel size

Fig. 2 Flow chart of included and excluded data


Patients infected with NTM or MTB with confirmed results and CT images (N=1351)
Tianjing Haihe Hospital, 01/2014-01/2020
MTB-LD (+) (N=952)
NTM-LD (+) (N=399)
Other pulmonary diseases (infectious diseases, tumor lesions, interstitial lesions) (N=49)
Other pulmonary diseases (infectious diseases, tumor lesions, interstitial lesions) (N=16)
MTB-LD (+) (N=903)
NTM-LD (+) (N=383)
Lung surgery (N=12)
Lung surgery (N=7)
MTB-LD (+) (N=891)
NTM-LD (+) (N=376)
Unqualified image (N=21)
Unqualified image (N=9)
(N=1237)
NTM-LD(+) MTB-LD(+) (N=66)
NTM-LD(+) MTB-LD(-) / NTM-LD(-) MTB-LD(+) (N=1105)
Center (N=1105)
Female : 310 Male : 795
NTM-LD : 301 MTB-LD : 804
Center2 (N=80)
XIAN Chest Hospital, 09/2019-03/2021
Training 80% (N=886)
Validation 10% (N=113)
Testing 10% (N=106)
External testing (N=80)
NTM-LD : 244 MTB-LD : 642
NTM-LD : 29 MTB-LD : 84
NTM-LD : 28 MTB-LD : 78
NTM-LD : 40 MTB-LD : 40

of Aquilion Prime 128 was 0.33–0.71 mm^3. The voxel sizes of all cases are shown in detail in supplementary material SS5.

Meanwhile, in order to compare the man–machine efficiency in the diagnosis of NTM-LD and MTB-LD, the results of the pathogenic microbiological examination were used to train the deep learning net, and an independent test set (n = 106) was randomly selected. Double-blind reading was employed by three radiologists with different experiences (senior experience: 10 years, reader 1; middle experience: 5 years, reader 2; junior experience: 3 years, reader 3) to determine whether the obtained CT images showed changes consistent with NTM-LD or MTB-LD. The average reading time of the learning model and radiologists for a single CT data was recorded.

Deep learning algorithm

The 3D-ResNet is a three dimensions modified version of the traditional two dimensions-residential network (2D-ResNet)[42], where 2D convolution is replaced by 3D convolution and pooling operation is replaced by 3D pooling, while the BatchNorm and activation function (ReLu) remained unchanged. The development of the 3D-ResNet-based NTM-LD/MTB-LD classifier in this study is shown in Fig. 3. Considering computation constraints, we normalised all CT data to a fixed scale of 64 × 256 × 256. For data preprocessing, the window width and level were adjusted to [1600, − 600], and the image intensity was normalised to [0, 1]. During training, 3D translation, rotation, Gaussian blur, and Gaussian noise were employed as data augmentation. To account for the smaller z-direction size of the input, we designed the 3D-ResNet to only have three stages, which consisted of four, six, and nine residual blocks, respectively. As a result, the feature map outputs by the final stage were down-sampled three times with a 1/8 resolution of the original input. For model training, the batch size was set to eight, and the initial learning rate was set to 0.1, which decayed by half after every 1000 iterations. We trained the model on four GeForce RTX 2080 Ti GPUs with the PyTorch framework for 10,000 iterations at maximum and would stop the training if the loss on the validation set failed to improve for 300 consecutive iterations.

Performance of the deep learning model

To contextualise the performance of the deep learning model, we invited three radiologists to blindly identify the pathogen that infected the patients in the testing dataset by

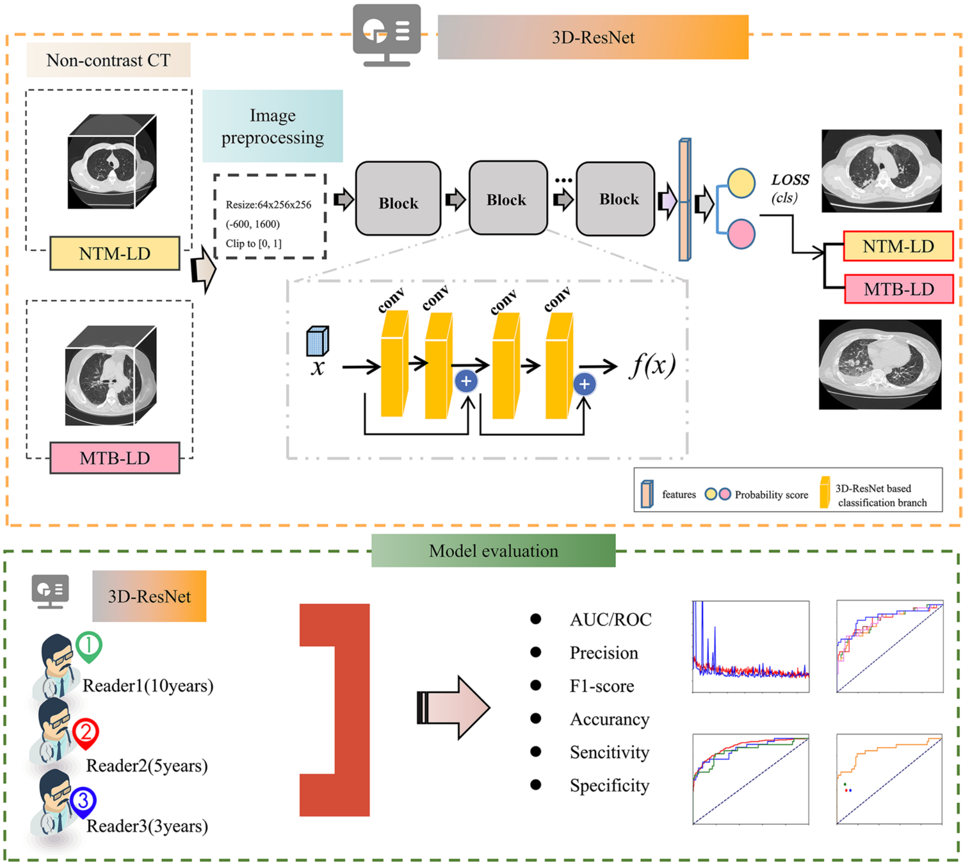


Fig. 3 Flow chart of the study. The first part was the training of 3D-ResNet, including 3D image preprocessing and 3D-ResNet construction of NTM-LD/MTB-LD classification network. The second part was the evaluation of deep learning model and man–machine comparison

only reading the CT scans. These readers were informed that the patients were only infected by one type of pathogen, either NTM or MTB. The reading time was also recorded for each radiologist during the interpretation of each CT scan.

To quantify the performance of the deep learning model and human experts, we computed different measures for accuracy and one measure for efficiency. The accuracy measures included area under the receiver operating curve (AUC), accuracy, sensitivity, specificity, precision, and F1 score. Efficiency was the average duration it took to classify each patient.

The deep learning model was an end-to-end predictive model, which directly learned the abstract mapping between lung CT images and resolving NTM-LD/MTB-LD. In order to demonstrate the reasoning process of the deep learning model more visually, we used the class activation map (CAM) to show the suspicious lung area detected by the model and reveal its potential value in making a diagnosis of NTM-LD and MTB-LD.

The network structure for 3D-ResNet is shown in Fig. 4, and it consists of four, six, and nine residual blocks, respectively.

Statistical analysis

SPSS version 22.0 (IBM Corp., Armonk, NY) was used to analyse the data. To analyse the difference between the clinical characteristics of the NTM-LD and the MTB-LD groups, an independent sample *t* test was used for numerical variables, and Pearson's chi-square analysis was used for categorical variables. This study was a two-tailed significance test, and *p*-values of < 0.05 were considered statistically significant. Python 3.0 (Python Software Foundation, https://www.python.org/) was used to draw receiver operating characteristic curve (ROC), confusion matrix, and data visualisation. The Delong test was used to analyse ROC.

Results

Basic clinical characteristics

A total of 1105 patients were included in this study, and this consisted of 301 patients in the NTM-LD group (age: 53.57 ± 17.21, 96 females, 205 males), and 804 patients in the MTB-LD group (age 47.11 ± 18.42, 214 females; 590 males). Table 1 displays the demographic characteristics of the two groups. Univariate analysis of 14 clinical characteristics using independent sample *t* test and Pearson's chi-squared test revealed that there among the demographic characteristics, patients with NTM-LD were older than those with MTB-LD ($p < 0.001$). Patients with MTB-LD had more cough ($p = 0.024$), and those with NTM-LD had more dyspnoea ($p = 0.011$), and the results were statistically significant.

Performance and evaluation of deep learning model

Using random assignment, the training set consisted of 886 patients (244 NTM-LD patients and 642 MTB-LD patients), the validation set had 113 patients (29 NTM-LD patients and 84 MTB-LD patients), and the testing set had 106 patients (28 NTM-LD patients and 78 MTB-LD patients). Finally, we had a training dataset of 18,080 $[(244 \times 2^5) + (642 \times 2^4)]$

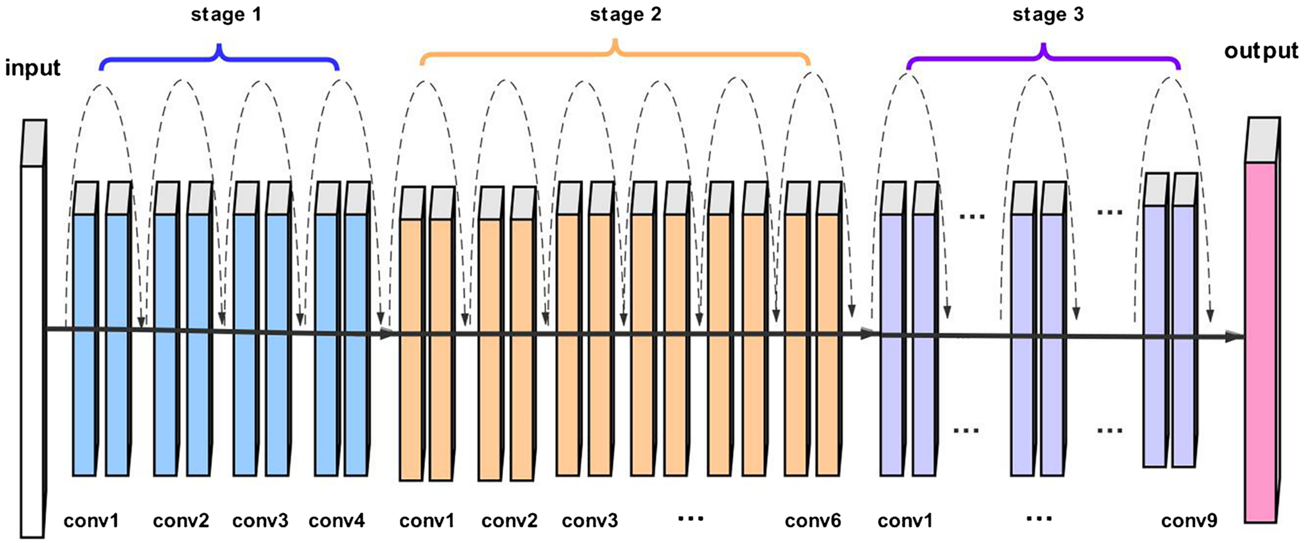


Fig. 4 Network structure of 3D-ResNet

images and a validation dataset of 2272 [$(29\times 2^5)+(84\times 2^4)$] images. The test set data remained unchanged without data augmentation.

The external test set contained 40 NTM-LD patients and 40 MTB-LD patients.

Training process and convergence of deep learning model

The loss curve of the model training is shown in Fig. 5a. The loss curve showed that when the iteration was at 0–4000 times, the model was under fitting; on the other hand, when the iteration was at 4000–6500 times, the model was in ideal training. Figure 5b shows the ROC and AUC of the classifier at different iterations on the validation set. The model trained on four GeForce RTX 2080 Ti GPUs with the PyTorch framework for 10,000 iterations, and the training time lasted for 3 h and 7 min. The 3D-ResNet trained for 6000 iterations was selected as the final classifier for testing.

Diagnostic performances of the deep learning model

The diagnostic performances of the selected classifier on the training, validation, and testing sets are shown in Table 2. The model fitted the training set well (AUC = 0.90, accuracy = 0.87, sensitivity = 0.98, specificity = 0.56) with good generalisability for the validation set (AUC = 0.88, accuracy = 0.88, sensitivity = 1, specificity = 0.52). When evaluated on the testing set, the overall model performed well (AUC = 0.86, accuracy = 0.83, sensitivity = 0.92, specificity = 0.57). In the external test set from Xi'an Chest Hospital, the model performed well (AUC = 0.78, accuracy = 0.69, sensitivity = 0.75, specificity = 0.63).

The ROCs of our model for the four datasets are uniformly plotted in Fig. 6. In Table 2, we list the performance of our model for the NTM-LD group and the MTB-LD group. We defined MTB-LD as a positive sample and NTM-LD as a negative sample.

Deep learning model identified suspicious lung areas

The CAM analysis demonstrated that our model after training was responsive to the suspicious lung areas that exhibited changes, including ground glass opacity, consolidation, nodule, tree-in-bud, and pleural effusion, which were also identified by the radiologists. Figure 7 shows the suspicious lung areas identified by the deep learning model in 3 NTM-LD patients (a–c) and 3 MTB-LD patients (d–f). Although the input region of interest (ROI) included both pulmonary and non-pulmonary areas, such as muscle and bone, without any manual annotation, the deep learning model was consistently focused on the inner lung region for prediction and was less affected by other tissue regions.

In addition, the deep learning model found that the suspected lung area overlapped highly with the actual inflammatory areas. Without manual annotation, the deep learning model could automatically focus on the changes in the bilateral lungs, such as consolidation, bronchiectasis, tree-in-bud, nodule, ground glass opacity, and fibrous cord, for identification and inference. We found high overlap and consistency between the radiologists and deep learning models. The results of the quantitative assessment of CAM and CT findings are presented in supplementary material SS2.

Table 1 Baseline characteristics of datasets

Clinical characteristics	All (n = 1105)	NTM-LD (n = 301)	MTB-LD (n = 804)	*p*-value
Age	48.87 ± 18.32	53.57 ± 17.21	47.11 ± 18.42	0.000[a]
Gender				0.082[b]
Female	310 (28.05%)	96 (31.89%)	214 (26.62%)	
Male	795 (71.95%)	205 (68.11%)	590 (73.38%)	
Chest pain	136 (12.31%)	39 (12.96%)	97 (12.06%)	0.779[b]
Cough	923 (83.53%)	234 (77.74%)	689 (85.70%)	0.024[b]
Expectoration	760 (68.78%)	203 (67.44%)	557 (69.28%)	0.702[b]
Fever	488 (44.16%)	118 (39.20%)	370 (46.02%)	0.099[b]
Chest tightness	109 (9.86%)	33 (10.96%)	76 (9.45%)	0.627[b]
Haemoptysis	163 (14.75%)	48 (15.95%)	115 (14.30%)	0.656[b]
Gasp	132 (11.95%)	48 (15.95%)	84 (10.45%)	0.080[b]
Dyspnoea	9 (0.81%)	7 (2.33%)	2 (0.25%)	0.011[b]
Chills	25 (2.26%)	5 (1.66%)	20 (2.49%)	0.474[b]
Fatigue	283 (25.61%)	91 (30.23%)	192 (23.88%)	0.132[b]
Night sweats	115 (10.41%)	28 (9.30%)	87 (10.82%)	0.614[b]
Weight loss	233 (21.09%)	69 (22.92%)	164 (20.40%)	0.511[b]

[a]Independent sample *t* test, [b]Pearson's chi-squared test.

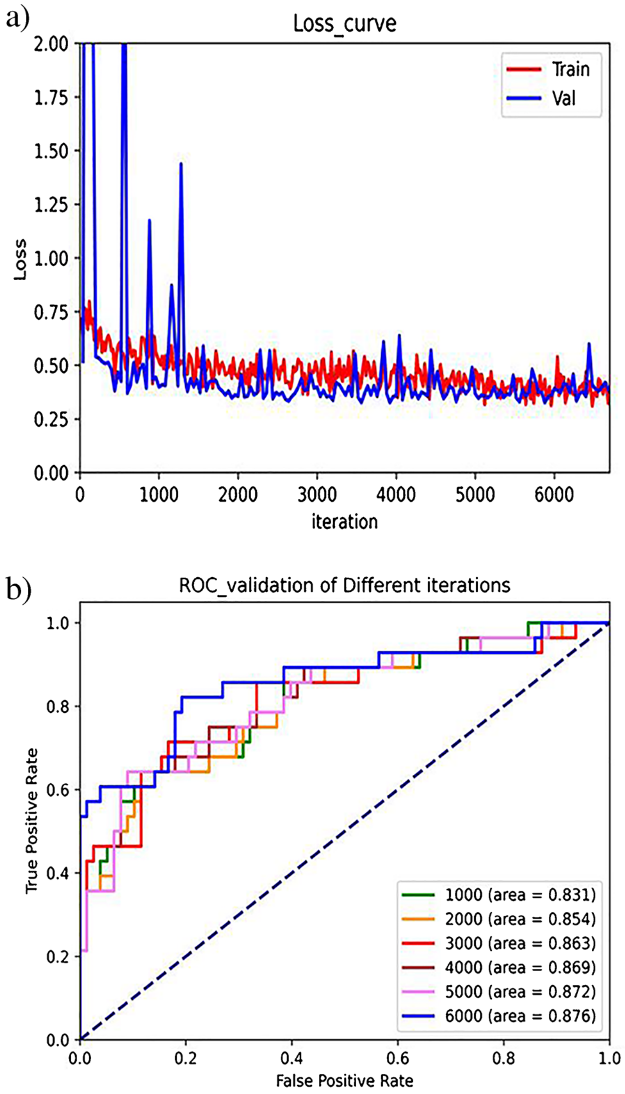


Fig. 5 **a** Loss curves of training set and validation set under different iterations. **b** ROC and AUC of the validation set under different iterations.

Comparison between man and machine

With the testing set, three radiologists with different experiences and the 3D-ResNet were used for man–machine comparison. The confusion matrix of the three radiologists and our model in distinguishing the NTM-LD from MTB-LD is shown in Fig. 8, and their diagnostic performance is detailed in Table 3. The discerning ability of the radiologists increased with experience, as shown by most accuracy metrics. For recalling NTM-LD, the 5-year experienced radiologist was not superior to the 3-year experienced radiologist (specificity = 0.39). The deep learning model performed better than all three radiologists in almost all the metrics, except for the recall of MTB-LD, where our model was equivalent to the senior radiologist (sensitivity = 0.92). The area under the ROC of our model was found to be significantly different from those of the three radiologists (Fig. 9) by the Delong test ($p < 0.05$). In addition, the deep learning model was significantly more efficient than the radiologists, requiring only 0.04 ~ 0.06 s/CT, which was nearly 1000 times faster than the radiologists.

Discussion

In this study, we developed a deep learning method based on chest CT images that showed good accuracy in distinguishing between NTM-LD and MTB-LD. The automatic activation area of the deep learning model was highly consistent with the area determined by radiologists, which suggests that deep learning automatically detects abnormalities in the lungs (consolidation, bronchiectasis, tree-in-bud, nodule, ground glass opacity, fibrous cord). This diagnostic method had the advantage of not requiring manual drawing ROI. Additionally, our model achieved a higher diagnostic accuracy with a higher efficiency rate than the radiologists.

This study collected the basic clinical datas and clinical manifestations of all patients included in the study. We found that patients with NTM-LD were older than those with MTB-LD. This is consistent with the findings of a previous study that found that NTM-LD was more common in the elderly than in young people [43]. Chiang et al. [44] showed that the proportion of females infected with NTM-LD was significantly higher than that of males. However, there is no difference in gender between NTM-LD and MTB-LD in this study, and this is consistent with the findings of Kim [26]. It was found that only two clinical symptoms, dyspnoea and cough, showed significant differences between the two groups. In this study, more patients with NTM-LD had dyspnoea compared to patients with MTB-LD, and this was consistent with the results of Klann [45]. However, Kim [30] reported that dyspnoea was more common in patients with MTB-LD. Therefore, we believe that patient characteristics and clinical manifestations alone are insufficient in differentiating between NTM-LD and MTB-LD.

Sensitivity and specificity served as important indices to evaluate the performance of the deep learning model and the radiologists to help improve the detection of NTM-LD and MTB-LD. Inverse relationships were observed between sensitivity and the number of misdiagnosed patients with MTB-LD, and between specificity and the number of

Table 2 Performance of 3D-ResNet on testing set, validation set, training set, and external test set

Performance	Testing set (n = 106) *NTM-LD (n = 28)* *MTB-LD (n = 78)*	Validation set (n = 113) *NTM-LD (n = 29)* *MTB-LD (n = 84)*	Training set (n = 886) *NTM-LD (n = 244)* *MTB-LD (n = 642)*	External test set (n = 80) *NTM-LD (n = 40)* *MTB-LD (n = 40)*
AUC (95% CI)	0.86 (0.76, 0.95)	0.88 (0.80, 0.96)	0.90 (0.88, 0.93)	0.78 (0.68, 0.89)
Accuracy	0.83	0.88	0.87	0.69
Specificity	0.57	0.52	0.56	0.63
Sensitivity	0.92	1	0.98	0.75
NTM-LD precision	0.73	1	0.93	0.71
MTB-LD precision	0.86	0.86	0.86	0.67
NTM-LD F1 score	0.64	0.68	0.70	0.67
MTB-LD F1 score	0.89	0.92	0.91	0.70

misdiagnosed patients with NTM-LD. Our model showed higher sensitivity and specificity than the radiologists, which indicated that our model had a lower misdiagnosis rate for NTM-LD and MTB-LD compared to that of the radiologists. Moreover, the model's diagnostic speed was > 1000 times faster than that of the radiologists, which can be considered an automatic rapid diagnosis. Deep learning models, when calibrated, would have the natural advantages of stability, repeatability, and ease of migration. As more chest CT images from other institutions are supplemented for training and testing, our model is expected to be applied to different institutions and would achieve better generalising capabilities.

According to previous studies, the use of deep learning models in the diagnosis of MTB has promising results. Lakhani [46]used 1007 chest radiographs from four datasets to construct deep convolutional neural networks (DCNNs) combined with AlexNet and GoogLeNet to classify patients as healthy or with tuberculosis infection (AUC = 0.99). Hwang [47] developed a deep learning automatic detection (DAD) algorithm using 5421 normal CXRs and 6768 active pulmonary MTB CXRs. The sensitivity and specificity of the algorithm were 94.3% and 91.1%, respectively, and the classification performance was significantly higher than that of the 15 physicians included in their study. Hwang [48]also designed a convolutional neural network-based (CNN-based) automated MTB screening computer-aided diagnosis (CAD) system based on many CXR images with transfer learning across three datasets. Qin [49] compared three deep learning (DL) systems (CAD4TB, Luna Insight, and QXR) based on CXRs to detect MTB in outpatients in Nepal and Cameroon. Seok [50] used 1000 CXR images to compare the performance of the I-CNN and the D-CNN from both image and demographic variables, and the DCNN was found to be better. Tawsifur [51]used multiple public databases to build a database containing 3500 MTB infections and 3500 normal chest radiographs that used nine different depth CNNs (RESNET18, RESNET50, RESNET101, CHEXNET, INCEPtionV3, VG19, DENSENET201, SQUEEZENET, and MobileNet) for transfer learning. The accuracy, precision, sensitivity, F1 score, and specificity of chest X-ray images were 96.47%, 96.62%, 96.47%, 96.47%, and 96.51%, respectively.

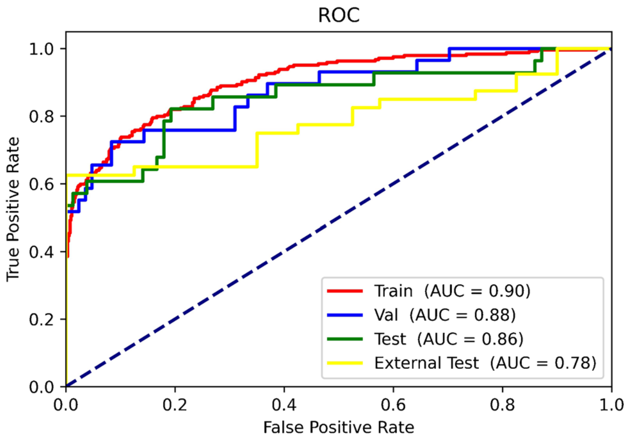


Fig. 6 The ROCs of 3D-ResNet on training set, validation set, test set, and external test set

There are some differences between our study and previous studies. First, our objective was to distinguish NTM-LD from MTB-LD based on CT images, since the CT manifestations of these two diseases were similar making differentiation difficult even for senior radiologists. Compared to previous studies, our objective was relatively more difficult; however, our model still performed excellently. Second, we compared our model to radiologists that

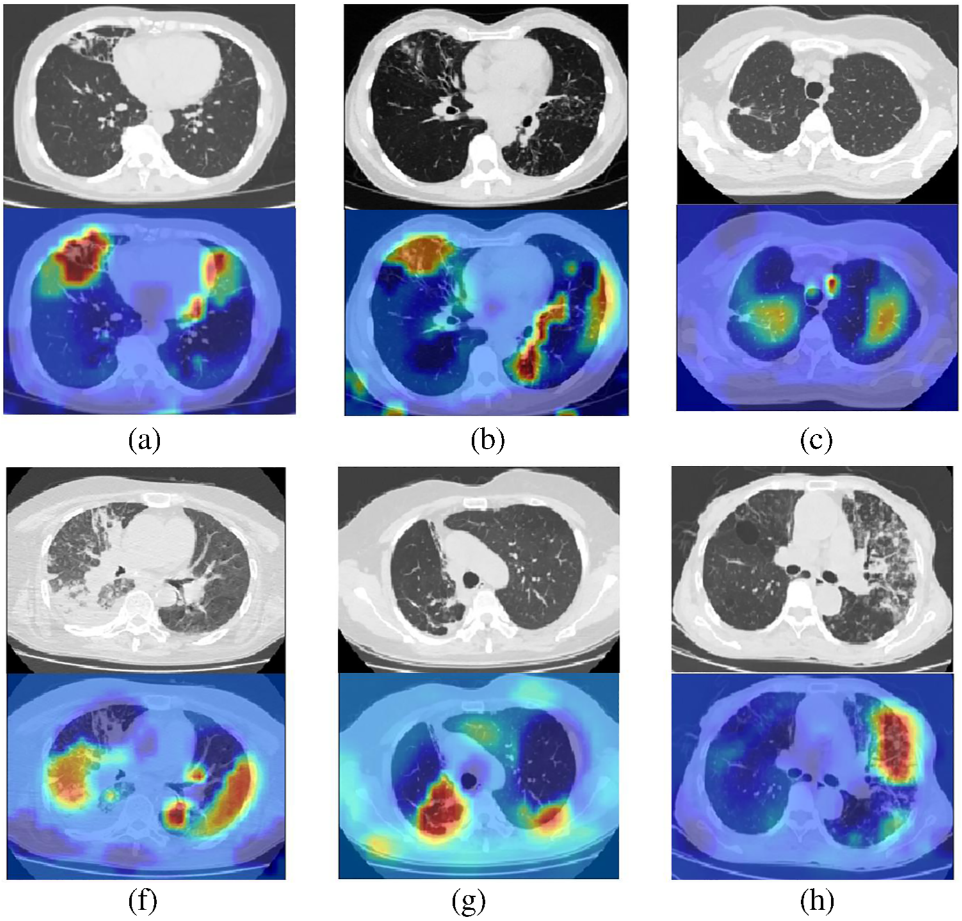


Fig. 7 CT images and class activation maps of 6 patients. **a** NTM-LD, consolidation with bronchiectasis in right middle lobe, fibrous cord in left upper lobe lingual segment. **b** NTM-LD, multiple bronchiectasis with central lobular nodules and tree-in-bud in right middle lobe and left lung. **c** NTM-LD, nodules and fibrous cord in right upper lobe, which sticked with pleura. **d** MTB-LD, consolidation in right upper lobe, right pleural effusion. **e** MTB-LD, consolidation and fibrous cord in right upper lobe, which sticked with pleura. **f** MTB-LD, multiple consolidation, nodules, tree-in-bud and ground glass opacity in left upper lobe

have 10, 5, and 3 years of experience on the test set, and the results showed that 3D-ResNet was significantly better than all radiologists in terms of accuracy, sensitivity, specificity, and AUC, which indicates good clinical applicability. Finally, we were able to easily collect 1105 chest CT images, suggesting that the model can be popularised and used in multiple institutions equipped with CT equipment. Additionally, the reference used for the diagnosis of the diseases was pathogenic microbiological examinations which were more time-consuming. The less time of diagnosis of our model further signifies the advantages of using the model.

Our study had several limitations. First, in the data collection, patients who were receiving specific drug therapy were not excluded. No study has shown whether clinical manifestations and chest CT findings are affected by specific drug therapy. However, the CT data of patients who received or did not receive specific drug treatment are prevalent in clinical scenarios, and this suggests that our model is suitable for practical medical scenarios. Second, the ratio of NTM-LD cases to MTB-LD cases was close to 3:8, suggesting that the sample was not sufficiently balanced. Although during the image preprocessing stage of model training, we used data resampling, data enhancement, and weight adjustment of loss function for optimisation, the performance of the model did not improve significantly ($p > 0.05$) (Supplementary material SS4 shows this in detail). This is directly related to the fact that the detection ability of the model in MTB-LD cases was far greater than that in NTM-LD cases (F1 score: 0.89 vs. 0.64). A more balanced sample ratio (1:1) will train the classifier to have a balanced detection ability. This also suggests that it is of definitive significance to supplement more NTM-LD cases in order to improve the model's ability to detect NTM-LD. Third, although the training, testing, and validation data were obtained from a single institution, we collected external test data for external validation of the model, and the model performed well; therefore, the addition of multicentre, multi-device, and multi-mode

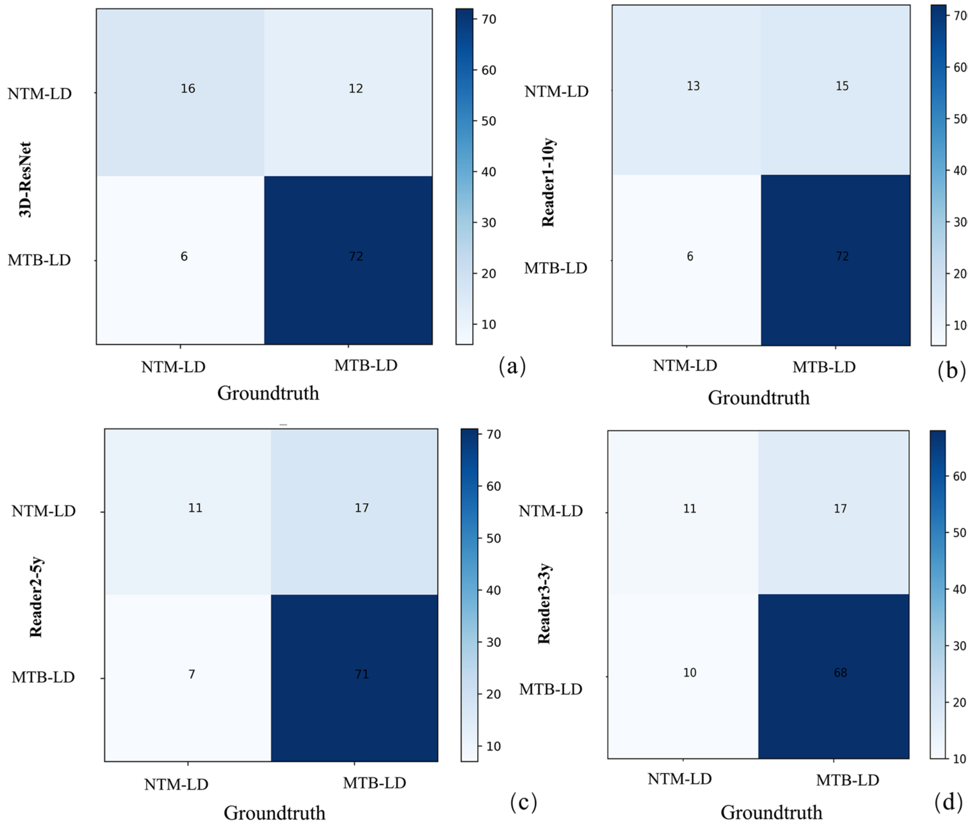


Fig. 8 Confusion matrix on the test set between the 3D-ResNet and three radiologists with different experiences (Reader1_10y, Reader2_5y, and Reader3_3y)

CT data will undoubtedly be of great significance to this study, which is also the direction of development of future work.

In conclusion, this study developed a personalised deep learning framework that uses conventional chest CT images to classify nontuberculous mycobacteria infection and *Mycobacterium tuberculosis* infection with greater accuracy, sensitivity, and specificity compared to those of radiologists. This model has the potential to be a screening tool for patients with pulmonary mycobacterium and will lead to early detection and selection of appropriate clinical treatment strategies.

Table 3 Comparison of discriminant performance between 3D-ResNet and 3 radiologists in the test set

Performance	3D-ResNet	Reader1-10y	Reader2-5y	Reader3-3y
AUC	0.86	0.69	0.65	0.63
Accuracy	0.83	0.80	0.77	0.75
Specificity	0.57	0.46	0.39	0.39
Sensitivity	0.92	0.92	0.91	0.87
NTM-LD_Precision	0.73	0.68	0.61	0.52
NTM-LD_F1 score	0.64	0.55	0.48	0.45
MTB-LD_Precision	0.86	0.83	0.81	0.80
MTB-LD_F1 score	0.89	0.87	0.86	0.83
sec/CT	0.04 ~ 0.06	58	62	71

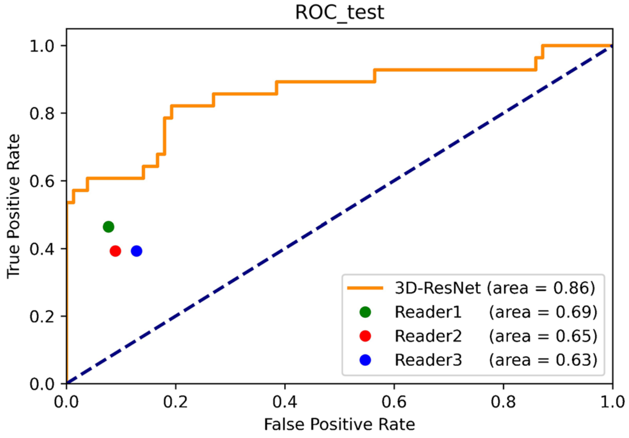


Fig. 9 ROC of 3D-ResNet and three radiologists (Reader1, Reader2, and Reader3) in the test set

Abbreviations *AI*, Artificial intelligence; *AUC*, Area under the curve; *CAD*, Computer-aided diagnosis; *CAM*, Class activation map; *CNN*, Convolutional neural network; *COVID-19*, Corona virus disease 2019; *CT*, Computed tomography; *CXR*, Chest X-ray; *DAD*, Deep learning automatic detection; *DL*, Deep learning; *DCNN*, Deep convolutional neural network; *MTB*, *Mycobacterium tuberculosis; MTB-LD*, *Mycobacterium tuberculosis* Lung disease; *NTM*, Nontuberculous mycobacterium; *NTM-LD*, Nontuberculous mycobacterium lung disease; *ROC*, Receiver operating characteristic curve; *ROI*, Region of interest; *ResNet*, Residential network; *2D*, Two dimensions; *3D*, Three dimensions

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00259-021-05432-x.

Acknowledgements We gratefully appreciate Xuan Wu and Yang Xu from Xi'an Chest Hospital for providing external test data.

Author contribution **Author list:** Li Wang, Wenlong Ding, Yan Mo, Dejun Shi, Shuo Zhang, Lingshan Zhong, Kai Wang, Jigang Wang, Chencui Huang, Shu Zhang, Zhaoxiang Ye, Jun Shen, Zhiheng Xing.

All authors contributed to the study conception and design. The research was designed by Zhiheng Xing, Chencui Huang, and Jun Shen. Material preparation and data collection were performed by Li Wang, Wenlong Ding, Shuo Zhang, Lingshan Zhong, and Kai Wang. Algorithm training and data analysis were performed by Yan Mo, Dejun Shi, and Li Wang. The first draft of the manuscript was written by Yan Mo and Li Wang and reviewed by Chencui Huang and Dejun Shi. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study has received funding by the Tianjin Municipal Health Science and Technology Project From Tianjin Municipal Health and Health Committee (2020xkm03) and Tianjin Haihe Hospital Project (HHYY-202007).

Data availability The CT images supporting the findings of this study are restricted by the Medical Ethics Committee of Haihe Hospital in order to protect patient privacy. If interested, requests for access to the extracted features can be made to the corresponding author Zhiheng Xing (18,920,696,025@189.cn).

Declarations

Ethics approval This retrospective study was approved by Haihe Hospital Ethics Committee (No. 2021HHKT-005). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

Consent to participate Yes.

Consent for publication Yes.

Conflict of interest The authors declare no competing interests.

References

1. World Health Organization. Global Tuberculosis Report 2020. https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf?sequence=1&isAllowed=y. Accessed 20 Jan 2021
2. Gopalaswamy R, Shanmugam S, Mondal R, Subbian S. Of tuberculosis and non-tuberculous mycobacterial infections - a comparative analysis of epidemiology, diagnosis and treatment. J Biomed Sci. 2020;27(1):74. https://doi.org/10.1186/s12929-020-00667-6.
3. Winthrop KL, Marras TK, Adjemian J, Zhang H, Wang P, Zhang Q. Incidence and prevalence of nontuberculous mycobacterial lung disease in a large United States managed care health plan, 2008–2015. Ann Am Thorac Soc. 2020;17(2):178–85.
4. Lee H, Myung W, Koh WJ, Moon SM, Jhn BW. Epidemiology of nontuberculous mycobacterial infection, South Korea, 2007–2016. Emerg Infect Dis. 2019;25(3):569–72.
5. Guidelines for the diagnosis and treatment of nontuberculous mycobacterial disease (2020) [J]. Chin J Tuberc Respir, 2020,43 (11): 918–946.
6. Yu X, Liu P, Liu G, Zhao L, Hu Y, Wei G, Luo J, Huang H. The prevalence of non-tuberculous mycobacterial infections in mainland China: systematic review and meta-analysis. J Infect. 2016;73(6):558–67. https://doi.org/10.1016/j.jinf.2016.08.020 .
7. Distribution and drug resistance of nontuberculous Mycobacteria in Beijing]. Zhonghua Jie He He Hu Xi Za Zhi. 2017 Mar 12;40(3):210–214. Chinese. doi: https://doi.org/10.3760/cma.j.issn.1001-0939.2017.03.013. PMID: 28297817.
8. Hong C, Li J, Zhao G, et al. Analysis on the epidemic situation of nontuberculous mycobacteria in Shenzhen from 2013 to 2017 [J]. Chin J Antituberc. 2019;41(5):529–33. https://doi.org/10.3969/j.issn.1000-6621.2019.05.011.
9. Zhu Yelei, AIzhen Pan, Lin Zhou, , et al. Analysis on epidemic situation and drug resistance of nontuberculosis mycobacterium in Zhejiang Province. Preventive medicine. 2021;33(01):6–10.
10. Yu Wang. Data compilation of the 5th national tuberculosis epidemiological sampling survey [M] Beijing: Military Medical Science Press, 2011:15–18.
11. Mathewos B, Kebede N, Kassa T, Mihret A, Getahun M. Characterization of mycobacterium isolates from pulmomary tuberculosis suspected cases visiting Tuberculosis Reference Laboratory at Ethiopian Health and Nutrition Research Institute, Addis Ababa Ethiopia: a cross sectional study. Asian Pac J Trop Med. 2015;8:35–40.
12. Cook JL. Nontuberculous mycobacteria: opportunistic environmental pathogens for predisposed hosts. Br Med Bull. 2010;96:45–59.
13. Miura K, Nakamura M, Taooka Y, Hotta T, Hamaguchi M, Okimoto T, Tsubata Y, Hamaguchi S, Kuraki T, Isobe T. Comparison of the chest computed tomography findings between patients with pulmonary tuberculosis and those with Mycobacterium avium complex lung disease. Respir Investig. 2020;58(3):137–43.
14. Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? Int J Tuberc Lung Dis. 2006;10(11):1192–204.
15. Reynolds J, Moyes RB, Breakwell DP. Appendix 3H: differential staining of bacteria—acid fast stain. In: CurrProtocMicrobiol 2009; appendix3: appendix 3H. Wiley Online Library website. http://onlinelibrary.wiley.com/doi/10.1002/9780471729259.mca03hs15/abstract. Accessed 2 Jan 2021.
16. Cowman S, Burns K, Benson S, Wilson R, Loebinger MR. The antimicrobial susceptibility of non-tuberculous mycobacteria. J Infect. 2016;72:324–31.

17. Campbell I, Drobniewski F, Novelli V, Ormerod P, Pozniak A. Management of opportunist mycobacterial infections: joint tuberculosis committee guidelines 1999. Thorax. 2000;55:210–8.
18. Lange C, Mori T. Advances in the diagnosis of tuberculosis. Respirology. 2010;15:220–40.
19 Kwak N, Lee CH, Lee HJ, Kang YA, Lee JH, Han SK, Yim JJ. Non-tuberculous mycobacterial lung disease: diagnosis based on computed tomography of the chest. Eur Radiol. 2016;26(12):4449–56.
20. Li L, Bannantine JP, Zhang Q, Amonsin A, May BJ, Alt D, Banerji N, Kanjilal S, Kapur V. The complete genome sequence of Mycobacterium avium subspecies paratuberculosis. Proc Natl Acad Sci USA. 2005;102:12344–9.
21. Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, Macheras M, Heym B, Herrman JL, Daffé M, Brosch R, Risler JL, Gaillard JL. Non mycobacterial virulence genes in the genome of the emerging pathogen Mycobacterium abscessus. PLoS ONE. 2009. https://doi.org/10.1371/journal.pone.0005660.
22 Koh WJ, Lee KS, Kwon OJ, Jeong YJ, Kwak SH, Kim TS. Bilateral bronchiectasis. and bronchiolitis at thin-section CT: diagnostic implications in nontuberculous mycobacterial pulmonary infection. Radiology. 2005;235:282–8.
23. Fujita J, Ohtsuki Y, Suemitsu I, Shigeto E, Yamadori I, Obayashi Y, Miyawaki H, DobashiN Matsushima T, Takahara J. Pathological and radiological changes in resected lung specimens in Mycobacterium avium intracellulare complex disease. Eur Respir J. 1999;13:535–40.
24. Hatipoglu ON, Osma E, Manisali M, Uc¸an ES, Balci P, Akkoclu A, Akpinar O, Karlikaya C, Yüksel C. . High resolution computed tomographic findings in pulmonary tuberculosis. Thorax. 1996;51:397–402.
25. Wu G, Chen X, Lin J, Wang Y, Yu J. Identification of invisible ischemic stroke in non-contrast CT based on novel two-stage convolutional neural network model. Med Phys. 2020. https://doi.org/10.1002/mp.14691.
26. Kim Y, Hahn S, Uh Y, Im D, Lim Y, Choi H, Kim HY. Comparable characteristics of tuberculous and non-tuberculous mycobacterial cavitary lung diseases. Int J Tuberc Lung Dis. 2014;18:725–9.
27. Li L, Chen Y, Bao Y, Jia X, Wang Y, Zuo T, Zhu F. Comparison of the performance between Frontier ASPECTS software and different levels of radiologists on assessing CT examinations of acute ischaemic stroke patients. Clin Radiol. 2020;75(5):358–65. https://doi.org/10.1016/j.crad.2019.12.010 .
28. Jeong YJ, Lee KS, Koh WJ, Han J, Kim TS, Kwon OJ. Nontuberculous mycobacterial pulmonary infection in immunocompetent patients: comparison of thin-section CT and histopathologic findings. Radiology. 2004;231:880–6. https://doi.org/10.1148/radiol.2313030833.
29. Chung MJ, Lee KS, Koh WJ, Kim TS, Kang EY, Kim SM, Kwon OJ, Kim S. Drug-sensitive tuberculosis, multidrug-resistant tuberculosis, and nontuberculous mycobacterial pulmonary disease in nonAIDS adults: comparisons of thin-section CT findings. Eur Radiol. 2006;16(9):1934–41. https://doi.org/10.1007/s00330-006-0174-9 .
30 Kim C, Park SH, Oh SY, Kim SS, Jo KW, Shim TS, Kim MY. Comparison of chest CT findings in nontuberculous mycobacterial diseases vs. Mycobacterium tuberculosis lung disease in HIV-negative patients with cavities. PLoS One. 2017;12(3):e0174240.
31. Yang C, Rangarajan A, Ranka S. Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification. AMIA AnnuSymp Proc. 2018 Dec 5;2018:1571–1580. PMID: 30815203.
32. Qayyum A, Ang C K, Sridevi S, Ahamed Khan MA, Hong LW, Mazher M, Chung TD. Hybrid 3D-ResNet deep learning model for automatic segmentation of thoracic organs at risk in CT images[C]//2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). IEEE, 2020:1–5.
33. Uemura T, Näppi JJ, Hironaka T, Kim H, Hiroyuki Y. Comparative performance of 3D-DenseNet, 3D-ResNet, and 3D-VGG models in polyp detection for CT colonography[C]//Medical Imaging 2020: Computer-Aided Diagnosis. International Society for Optics and Photonics. 2020;11314:1131435.
34. Zhou T, Lu H, Yang Z, Qiu S, Huo B, Dong Y. The ensemble deep learning model for novel COVID-19 on CT images. Appl Soft Comput. 2020;6:106885.
35. Keles A, Keles MB, Keles A. COV19-CNNet and COV19-ResNet: diagnostic inference engines for early detection of COVID-19. Cognit Comput. 2021;6:1–11.
36. Yang Y, Zhang Q. Multiview framework using a 3D residual network for pulmonary micronodule malignancy risk classification. Biomed Mater Eng. 2020;31(4):253–67.
37. Zhang F, Wang Q, Li H. Automatic segmentation of the gross target volume in non-small cell lung cancer using a modified version of resNet. Technol Cancer Res Treat. 2020;19:1533033820947484.
38. Ning J, Zhao H, Lan L, Sun P, Feng Y. A computer-aided detection system for the detection of lung nodules based on 3D-ResNet. Appl Sci. 2019;9(24):5544.
39. Polat H, DanaeiMehr H. Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture. Appl Sci. 2019;9(5):940.
40. Daley CL, Iaccarino JM, Lange C, Cambau E, Wallace RJ Jr, Andrejak C, Böttger EC, Brozek J, Griffith DE, Guglielmetti L, Huitt GA, Knight SL, Leitman P, Marras TK, Olivier KN, Santin M, Stout JE, Tortoli E, van Ingen J, Wagner D, Winthrop KL. Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. Eur Respir J. 2020;56(1):2000535.
41. National Health Commission of the People's Republic of China. In: Diagnostic criteria. for pulmonary tuberculosis(WS 288-2017) [EB/OL].(2017-11-20) [2020-04-10]. 2017. http://www.nhc.gov.cn/ewebeditor/uploadfile/2017/11/20171128164254246.pdf. Accessed 4 Jan 2021.
42. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Computer vision and pattern recognition. Cornell University. 2016. https://arxiv.org/abs/1603.05027. Accessed 4 Jan 2021.
43. Winthrop KL, McNelley E, Kendall B, Marshall-Olson A, Morris C, Cassidy M, Saulson A, Hedberg K. Pulmonary non-tuberculous mycobacterial disease prevalence and clinical features: an emerging public health disease. Am J Respir Crit Care Med. 2010;182(7):977–82.
44 Chiang CY, Yu MC, Yang SL, Yen MY, Bai KJ. Surveillance of tuberculosis in Taipei: the influence of nontuberculous Mycobacteria. PLoS One. 2015;10(11):e0142324e.
45. Klann E, Beal SG, Tremblay EE. Evaluating differences in tuberculosis and nontuberculous mycobacterial lung disease in Florida. Am J Infect Control. 2019;47(11):1324–8.
46. 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.
47. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, Goo JM, Aum J, Yim JJ, Cohen J, Ferretti GR, Park CM. Development and validation of a deep learning–based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. Clin Infect Dis. 2019;69(5):739–47.
48. Hwang S, Kim HE, Jeong J, Kim HJ. A novel approach for tuberculosis screening based on deep convolutional neural networks[C]//

Medical imaging 2016: computer-aided diagnosis. International Society for Optics and Photonics. 2016;9785:97852W.

49. Qin ZZ, Sander MS, Rai B, Titahong CN, Sudrungot S, Laah SN, Adhihkari LM, Carter EJ, Puri L, Codlin AJ, Creswell J. Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. Sci Rep. 2019;9(1):1–10.

50. Heo SJ, Kim Y, Yun S, Lim SS, Kim J, Nam CM, Park EC, Jung I, Yoon JH. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. Int J Environ Res Public Health. 2019;16(2):250e.

51. Rahman T, Khandakar A, Kadir MA, Islam KR, Islam KF, Mazhar R, Hamid T, Islam MT, Mahbub ZB, Ayari MA, Chowdhury MEH. Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. IEEE Access. 2020;8:191586–601.

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