European Journal of Radiology 169 (2023) 111180

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Deep learning and radiomics of longitudinal CT scans for early prediction of tuberculosis treatment outcomes

Mayidili Nijiati [a,1], Lin Guo [b,1], Abudoukeyoumujiang Abulizi [a,1], Shiyu Fan [a], Abulikemu Wubuli [c], Abudouresuli Tuersun [a], Pahatijiang Nijiati [a], Li Xia [b], Kunlei Hong [b], Xiaoguang Zou [d,*]

[a] Department of Radiology, The First People's Hospital of Kashi (Kashgar) Prefecture, China
[b] Shenzhen Zhiying Medical Imaging, Shenzhen, China
[c] Department of Radiology, Yecheng County People's Hospital, China
[d] Clinical Medical Research Center, The First People's Hospital of Kashi (Kashgar) Prefecture, China

ARTICLE INFO

Keywords:
Longitudinal CT scans
Radiomics
Deep learning
Treatment outcome prediction
Drug-resistant tuberculosis

ABSTRACT

Background: To predict tuberculosis (TB) treatment outcomes at an early stage, prevent poor outcomes of drug-resistant tuberculosis (DR-TB) and interrupt transmission.
Methods: An internal cohort for model development consists of 204 bacteriologically-confirmed TB patients who completed anti-tuberculosis treatment, with one pretreatment and two follow-up CT images (612 scans). Three radiomics feature-based models (RM) with multiple classifiers of Bagging, Random forest and Gradient boosting and two deep-learning-based models (i.e., supervised deep-learning model, SDLM; weakly supervised deep-learning model, WSDLM) are developed independently. Prediction scores of RM and deep-learning models with respectively highest performance are fused to create new fusion models under different fusion strategies. An additional independent validation was conducted on the external cohort comprising 80 patients (160 scans).
Results: For RM scheme, 16 optimal radiomics features are finally selected using longitudinal scans. The AUCs of RM for Bagging, Random forest and Gradient boosting were 0.789, 0.773 and 0.764 in the internal cohort and 0.840, 0.834 and 0.816 in the external cohort, respectively. For deep learning-based scheme, AUCs of SDLM and WSDLM were 0.767 and 0.661 in the internal cohort, and 0.823 and 0.651 in the external. The fusion model yields AUCs from 0.767 to 0.802 in the internal cohort, and from 0.831 to 0.857 in the external cohort.
Conclusions: Fusion of radiomics features and deep-learning model may have the potential to predict early failure outcome of DR-TB, which may be combined to help prevent poor TB treatment outcomes.

1. Introduction

Tuberculosis (TB) is one of the deadliest infectious diseases, and WHO reported that the COVID-19 pandemic has a damaging impact on access to TB diagnosis and treatment. TB treatment and control actions are urgently needed [1]. The treatment therapy for TB varies depending on each patient's type and individualized level, and drug resistance is associated with poorer outcomes or treatment failures [2]. Drug-resistant TB (DR-TB) is recognized as being resistant to the two most effective first-line drugs of isoniazid and rifampicin; thus the treatment success rate of DR-TB is lower than drug-sensitive TB but with a much higher cost [3,4]. DR-TB is one of the major obstacles in the global fight against tuberculosis because it may complicate therapy and raises the probability of poor results. The early prediction of drug-resistant tuberculosis (TB) has the potential to significantly impact TB control efforts and improve patient outcomes,for example, timely treatment initiation, targeted interventions, improved treatment outcomes, reduced transmission, efficient resource allocation, and surveillance, monitoring [5]. It is especially of great importance in poor, resource limited areas. It has been reported that clinical risk factors of the number of earlier treatments and the score of sputum smears tests would help predict DR-TB [6]. However, additional clinical information and pathological evaluations are not always available, especially in resource-limited rural areas. Demographics and clinical parameters such as age,

* Corresponding author at: The First People's Hospital of Kashi (Kashgar) Prefecture, No. 120, Yingbin Avenue, Kashi (Kashgar), Xinjiang, China.
E-mail address: zxgks@163.com (X. Zou).
[1] Mayidili Nijiati, Lin Guo and Abudoukeyoumujiang Abulizi contributed equally to this work.

https://doi.org/10.1016/j.ejrad.2023.111180
Received 25 July 2023; Received in revised form 21 October 2023; Accepted 29 October 2023
Available online 30 October 2023

gender and regimen of the drug play an essential part in predicting the TB treatment outcome [7,8] but not taking into account phenotypic changes of TB. It has been reported that methods of genome sequencing and nucleic acid amplification are also used to detect DR-TB [9,10]. However, they are both too costly to be applied in primary medical settings with limited resources.

CT imaging plays a vital role in monitoring the patients treatment process [11–14], and recently radiomics and deep learning (DL) technique have been demonstrated as potential tools to automatically detect TB using static CT images over an individual timepoint [15,16]. However, few of them applied both radiomics and DL to create a fusion model to predict TB treatment outcomes. Furthermore, to the best of our knowledge, we are also the first study to evaluate whether longitudinal CT-based radiomics features and DL network could predict DR-TB. In the study, to predict the TB treatment outcomes of success and DR-TB, we first developed three radiomics models (RM) by using multiple classifiers, and two DL-based models (i.e., supervised deep learning model, SDLM; weakly supervised deep learning model, WSDLM) using longitudinal CT scans. Then, we have fused the radiomics feature based model and DL model with the highest performance to build a new AI scheme to predict treatment outcomes between DR-TB and success. In order to evaluate the performance of different schemes, all the established models are independently validated on an external cohort.

2. Methods

2.1. Patient cohorts

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), and it was approved by the local hospital Institutional Review Board with a waiver of informed consent of patients for the retrospective research. TB Patients who are bacteriologically confirmed and have already completed treatment were collected from two local hospitals in China, named Internal cohort and External cohort. According to the inclusion and exclusion criteria, a total of 284 eligible patients were finally included from the original 5255 patients (Fig. 1). Internal cohort, comprising 204 patients from Site A, was collected between Jan 1, 2018 and June 30, 2021 with training, validation and testing at a ratio of 7:1:2. External cohort was subsequently collected between June 1, 2019 and June 30, 2021 form Site B, consisting of 80 patients.

The criteria for exclusion were determined as (i) TB patients confirmed by Acid-fast bacilli culture positivity with presentence of M. tuberculosis; (ii) TB patients completed the anti-tuberculosis treatment in the local hospitals, and all the clinical, laboratory and imaging information was acquired; (iii) Each patient had three times CT scans, one is the pretreatment CT scan, and the other two were follow-up CT scans taken at the end of the second month and the sixth month (end of treatment); (iv) Three types of treatment outcomes were included: cured, treatment completed and DR-TB. The complete exclusion flow chart is provided in Fig. 1.

In this study, 284 eligible patients received a standard 6-month 2HRZ/4HR regimen, using a 2-month intensive phase of daily isoniazid (INH), rifampicin (RMP), and pyrazinamide (PZA), followed by a 4-month continuation phase of daily INH and RMP. At the end of the second month of the regimen, patients underwent a CT scan and sputum culture examination. If sputum culture remained positive, drug susceptibility testing was conducted. If the patient was found to be drug-sensitive, the regimen remained unchanged, but an additional sputum examination was performed at the end of the third month. If the patient was drug-resistant, the regimen was extended to 12–18 months, and second-line anti-TB drugs were prescribed after the 12-to-18-month regimen. At the end of the original treatment regimen, which is the end of the sixth month, all patients underwent a CT scan. The 6-month treatment outcomes were classified according to the guideline of WHO [17], including (i) Cured, TB patients who completed the 6-month regimen with bacteriological confirmation of the initial diagnosis of TB and who have negative bacteriological results (smear-or-culture-negative) on at least two occasions, one of which should be at the end of treatment; (ii) Treatment completed, TB patients who completed the 6-month regimen without evidence of failure but with no record to show that sputum smear or culture results on at least two occasions; (iii) DR-TB, TB patients who were initially diagnosed with drug-susceptible TB but later developed drug resistance during the course of treatment (commonly referred to as acquired drug resistance), including multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB); (iv) Treatment success, the sum of cured and treatment completed. The internal cohort contained 160 success patients and 44 DR-TB patients, and the external cohort contained 65 success patients and 15 DR-TB patients.

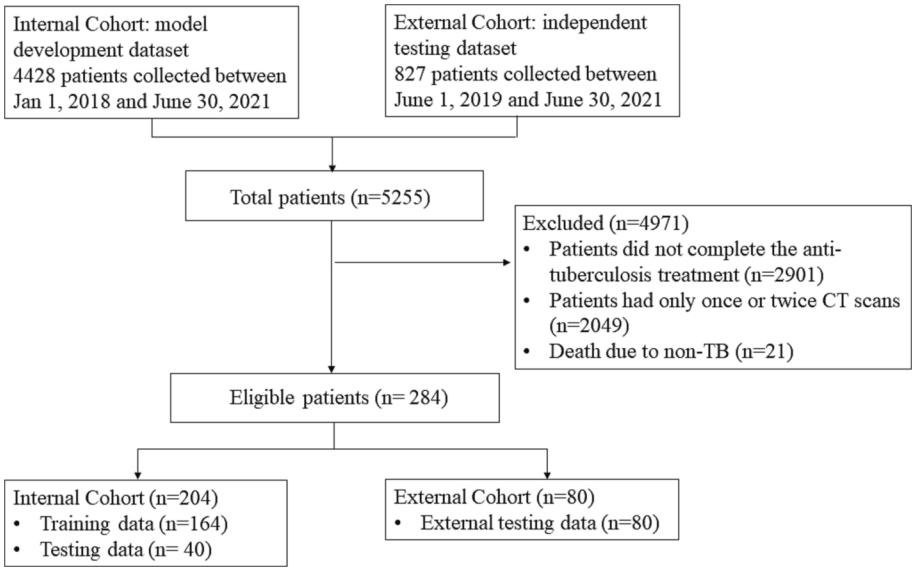

Fig. 1. Flowchart for study dataset. A total of 5255 patients from the multicenter were initially involved in the investigation, and 284 patients were finally included of which 204 patients were used as the internal dataset, and 80 patients were for the external validation.

2.2. CT image acquisition and image preprocessing

CT scans were acquired according to standardized protocols at each hospital, and all identification information of the patients was removed before the imaging preprocessing. CT scans in the internal cohort were performed using United Imaging 16-row 32-slice helical CT scan (tube voltage: 120 kV, tube current: 100 mA, pitch: 1.5, slice thickness: 7.0 mm, field of view: 450 mm) and PHILIPS Brilliance 32-row helical CT scan (tube voltage: 120 kV, tube current: 100 mA, pitch: 1.5, slice thickness: 5.0 mm, field of view: 450 mm). CT scans in the external cohort were performed using Siemens 64-row 128-slice helical CT scan (SOMATOM Definition AS, tube voltage: 100 kV, tube current: 100 mA, pitch: 1.3, slice thickness: 5.0 mm, field of view (FOV): 430 mm).

Two radiologists (more than 10 years of experience in reading CT images) outlined TB lesions on each slice of CT scans to produce a Dice coefficient value individually. If two Dice coefficient values were all greater than or at least equal to 0.95, they would be averaged as the ground truth of the image. Otherwise, a senior radiologist (more than 20 years of experience in reading CT images) would review and outline the images again to make the final determination. Image augmentation, including image flipping, translation, rotation and deformation was performed on the internal cohort to improve the model performance using established methods [18].

2.3. Development of RM

As shown in Fig. 2, a variety of filtering processes, including squaring, square root, logarithm, exponential, gradient, wavelet transform operations, etc., were performed on the CT scans with manual lesion contours. Then a total of 107 features were obtained, including 18 first order features, 14 shape features and 75 texture features. We extracted those radiomics features and normalized them to [0, 1]. To reduce the dimensionality of extracted features, we applied LASSO method to select the most predictive features to develop RM. Finally, 16 selected features were used to train classifiers of Bagging, Random forest (RF) and Gradient boosting (Gboost) to develop multiple RMs.

2.4. Development of SDLM and WSDLM

As shown in Fig. 2, a 3D-dimensional central neural network was used to develop SDLM and WSDLM. For the SDLM, the input to this network was the region of interest (ROI) of TB, and the 3D Resnet was used to extract features. Then the Gated Recurrent Unit (GRU), which consisted of 2-layer recurrent neural network units and one fully layer was applied to make the predictions, during which ROI features extracted at a different time of the treatment period needed to input to GRU in sequence. The WSDLM was developed in Python with the Pytorch backend (Python 3.8, Pytorch 1.10.1). Each slice of the CT image was reduced to 320 from 512, then combined into 3D images. The input into the 3D Resnet in WSDLM scheme was all time series CT scans of a patient without manual lesion contours instead of the ROIs in SDLM scheme, to extract features and obtain the feature maps. To capture the interval feature changes, feature maps of the next time were aligned to the previous one to make subtractions from each other. Finally, multi-scale average pooling was performed to obtain the fusion features, and global average pooling and fully connected layers were applied to make classifications of success and DR-TB cases. Therefore, the WSDLM predicts by automated feature map alignment and subtraction of longitudinal scans, which differs from the SDLM with manual delineations for DL model training.

2.5. Model fusion

After training and building the three RMs and two DL models, we compared the performance of the trained models and selected one RM and one DL model with the highest AUC values respectively as the optimal one. The range of prediction scores generated from radiomics feature-based scheme and deep learning-based scheme is distributed in different scales either on internal or external testing; therefore data preprocessing procedure of normalization is necessary before model fusion to bring all the attributes on the same scale. The original optimal operating points for all the cases are set to be equal to 0.5, and the prediction scores are accordingly normalized to the same standard scale [0,1]. Finally, the RM using Bagging and SDLM were fused by weighting their prediction scores. For the weighting average strategy, we constantly increased the weighting factor of the SDLM prediction score from 0 to 1 with an interval of 0.1. The weighting factor of RM correspondingly decreased from 1 to 0 to compute the fusion prediction score.

2.6. Model performance evaluation

Statistical analysis was performed using Python 3.8 and SPSS 20. The

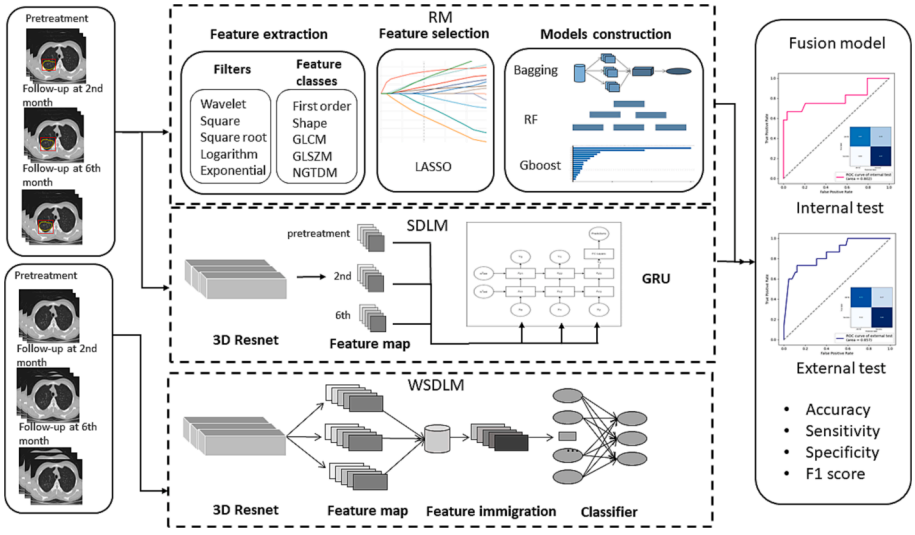

Fig. 2. Illustration of different models' architecture. Multiple TB treatment outcome prediction models of three RMs (i.e., multiple classifiers of Bagging, Random forest and Gradient boosting) and two DL models (i.e., SDLM, WSDLM)) were developed separately using longitudinal CT scans, and the RM with the highest AUC value as the optimal one was combined with the optimal DL model to create a fusion model. The performance of the fusion model was tested on both internal and external cohorts. TB, tuberculosis; RM, radiomics model; DL, deep learning.

performance of the proposed models to predict treatment outcomes on the internal and external cohort were assessed by the receiver operating characteristic curve (ROC) and area under the curve (AUC). Besides, the accuracy, sensitivity, specificity, and F1 score were also calculated and compared.

3. Results

3.1. Clinical characteristics

The main characteristics of patients in the internal and external cohorts are shown in Table 1. The two cohorts were inherently different. The internal cohort was used to train and test the RM and DL models, and the external cohort was used to independently test the model performance. 204 patients from the internal cohort were 55.9 % males, and the median age was 67 years, with an age range of 14–93 years. The external cohort included 80 patients with 58.8 % males (median age of 65.5 years; age range 18–87 years) (Fig. S1). It is not statistically significant in the patient age ($P = 0.388$), gender ($P = 0.334$) for both two cohorts, neither for the outcome distribution of success and DR-TB ($P = 0.598$), but there is a significant difference in the groups of cured, treatment completed and transferred to drug-resistant ($P = 0.002$).

3.2. Radiomics feature selection

There were 107 radiomics features extracted from the initial feature pool, and not all features contribute to the positive performance of treatment outcome prediction, and some features might add noise to it. The LASSO model is introduced in this study to estimate what variables are important in the prediction. The parameter λ for the LASSO model was set to 0.01 based on the cross-validation. The 16 most important predictive biomarkers were finally identified for model construction. The normalized importance of the 16 features is shown in Fig. 3. There were 3 shape features, 9 Gy level co-occurrence matrix texture features (GLCM), a gray level run length matrix texture feature (GLRLM), 2 Gy level size zone matrix texture features (GLSZM), and a neighborhood gray-tone difference matrix (NGTDM).

3.3. Comparison of multiple RMs and DL models

After radiomics feature extraction and selection, they were used for different RMs training, and the comparisons of Bagging, RF and Gboost in the AUCs were presented in Fig. 4a, b. For the internal cohort, Bagging, RF and Gboost obtained AUCs of 0.789, 0.773 and 0.764, respectively. While in the external cohort, the AUCs were 0.840, 0.834 and 0.816. We observed that about 70 % DR-TB and over 80 % success patients could be predicted correctly (Fig. 4c, d). In both internal and external cohorts, the prediction performance of Bagging outperformed RF and Gboost, with Gboost as the poorest one. Therefore, Bagging was selected as the optimal RM to be involved in creating the new fusion model. As shown in Fig. 5, the SDLM obtained AUCs of 0.767 and 0.823 in the internal and external cohorts, which outperformed the WSDLM with AUCs of 0.661 and 0.651, respectively (Fig. 5a, b). The SDLM identified 70 % DR-TB and over 80 % success patients in the external cohort, but only 60 % DR-TB and success patients were predicted in the internal cohort (Fig. 5c, d). However, it is consistent with the performance of RM. Finally, SDLM was selected as the optimal DL model and incorporated with Bagging to create the new fusion model.

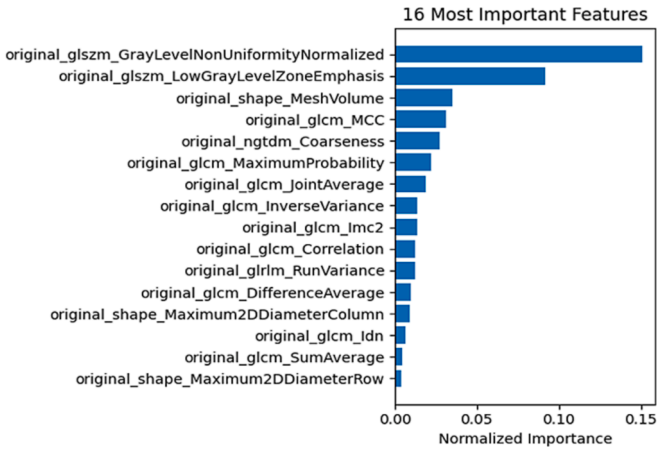

Fig. 3. The 16 most predictive features were selected and used for RM development. RM, radiomics model.

3.4. Performance evaluation of Bagging, SDLM and fusion model

The accuracy, sensitivity, specificity and F1 score are shown in Table 2. For the Bagging, the accuracy, sensitivity, specificity and F1 score were 0.805 vs. 0.800, 0.667 vs. 0.733, 0.862 vs. 0.815 and 0.667 vs. 0.579 in the internal and external cohort, respectively. While for the WSDLM, the accuracy, sensitivity, specificity and F1 score were 0.585, 0.583, 0.586 and 0.452 in the internal cohort, which was confirmed in the external cohort of 0.875, 0.733, 0.908 and 0.688. Results indicated that the optimal RM of Bagging outperformed the SDLM on both internal and external cohorts. Table 3 lists the AUC values and the corresponding 95 % confidence interval (CI) of the fusion models proposed in this study. Testing on internal and external cohorts, the RM-based scheme of Bagging and the DL-based scheme of SDLM yielded AUC values of 0.789 vs. 0.767 and 0.840 vs. 0.823, respectively. The new scheme performance changed with the different fusion strategies when we applied the fusion method. The new final fusion model under different fusion strategies yielded AUCs from 0.767 to 802 in the internal cohort, and from 0.831 to 0.857 in the external cohort (Table 3). Using the maximum fusion strategy, the fusion model could achieve the highest AUC values of 0.802 (95 % CI, 0.648 to 0.910) in the internal cohort and 0.857 (95 % CI, 0.761 to 0.926) in the external cohort (Fig. 6). In addition, the accuracy, sensitivity, specificity and F1 score of fusion scheme with optimal performance (i.e., internal AUC = 0.802; external AUC = 0.857) were 0.878, 0.667, 0.966 and 0.762 in the internal cohort, and 0.850, 0.733, 0.877 and 0.647 in the external cohort (Table 4).

Table 1
Patient demographics and outcomes of 2 different cohorts.

Demographics	Internal Cohort (n = 204)	External Cohort (n = 80)	T/χ^2	*P* Value
Age (yr, mean ± sd.)	61.24 ± 18.42	59.14 ± 18.29	0.865	0.388
Sex (n male)	221 (52.5 %)	100 (56.8 %)	0.934	0.334
Outcomes				
Success (n (%))	160 (78.4 %)	65 (81.2 %)	0.277	0.598[a]
-Cured (n (%))	88 (41.1 %)	52 (65.0 %)	12.755	0.002[b]
-Treatment completed (n (%))	72 (35.3 %)	13 (16.2 %)	NA	NA
Transferred to drug-resistant treatment (n (%))	44 (21.6 %)	15(18.8 %)	NA	NA

[a] Comparison of outcomes of success and transferred to drug-resistant.
[b] Comparison of outcomes of cured, treatment completed and transferred to drug-resistant.
Abbreviations: NA, not applicable.

4. Discussion

In this study, we developed three radiomics models (i.e., Bagging, RF and Gboost), two deep learning models (SDLM and WSDLM), and a fusion scheme with the different fusion strategies to predict TB treatment outcomes using longitudinal CT images in both internal and external cohorts.

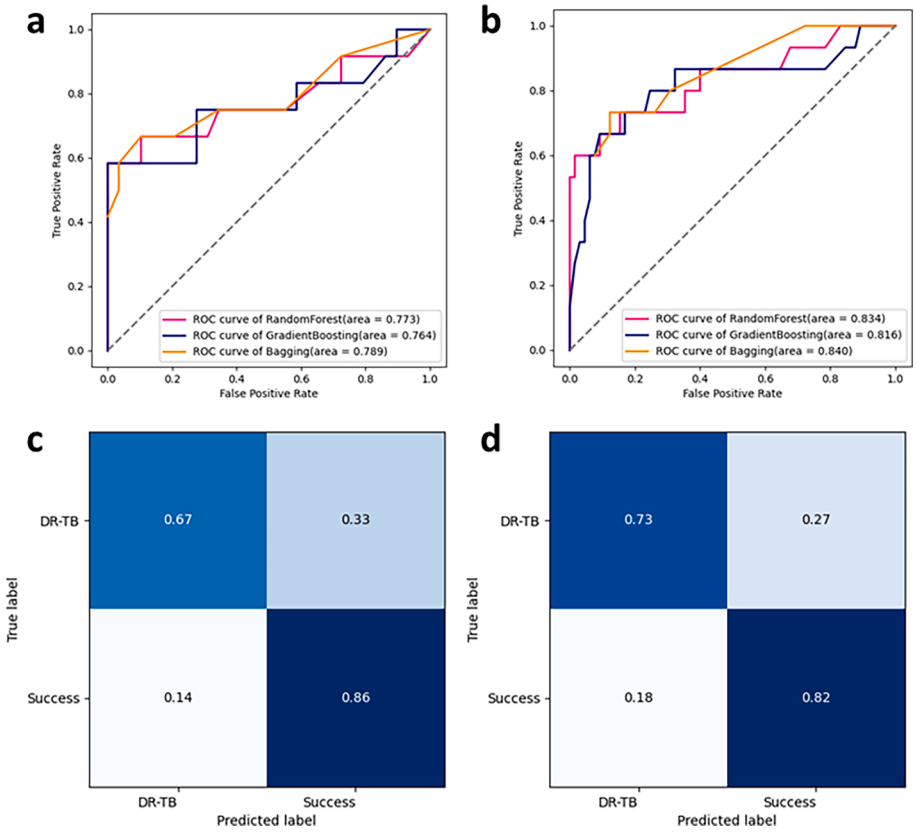

Fig. 4. Performance of the radiomics feature-based models in predicting TB treatment outcome. (a) ROC curves of three different radiomics models in the internal cohort; (b) ROC curves of three different radiomics models in the external cohort; (c) Normalized confusion matrix of the bagging model in the internal cohort; (d) Normalized confusion matrix of the bagging model in the external cohort. TB, tuberculosis; ROC, receiver operating characteristic.

Clinical and demographic features were often used to predict treatment outcomes, and the number of previous treatments, lack of a job and alcohol consumption were reported to be potential risk factors for the occurrence of DR-TB with an AUC of 0.74 [19]. However, such clinical and demographic information does not consider the phenotypic changes. More importantly, different factors were reported to have the strongest predictive performance, making it hard to obtain a general factor for practical application in clinical settings [20,21,7].

Radiomics is an effective tool to detect TB from CT images as it can offer internal features such as texture features, wavelet features and histogram features except for common features of shape and volume, which would help reflect more detailed information for TB lesions. In the study, 3 shape features and 13 texture features are the most predictive biomarkers to differentiate DR-TB. The shape feature describes the geometry of the ROI, which quantifies the shape of the ROI to reflect its degree of sphericity, and the texture feature represents the intensity level of the spatial distribution of voxels. Image texture is a spatial change that can be perceived or measured at the intensity level and is often considered a gray scale. The results showed that the DR-TB group differs from the success group in textures, which is consistent with the findings that texture features are relevant to DR-TB [16]). Previous studies have reported that radiomics model could distinguish TB from lung cancer using CT scans [22–24], but very few focused on DR-TB classification; a recent study reported a good performance was achieved to detect DR-TB from drug-sensitive TB with an AUC of 0.844 and 0.829 in the training cohort and testing cohort, respectively [16]. The proposed radiomics feature based scheme in our study obtained a comparable performance with AUC values of 0.789 and 0.840 in internal and external cohorts. In our research, the final follow-up CT scans at the end of treatment were not involved in the testing to investigate the potential for prediction at an early stage, making it challenging for the radiomics tool to identify DR-TB.

Various DL-based approaches to automated screening and detection of tuberculosis from radiological images achieved promising results [15,25,26]. In the current study, we applied different supervised and weakly supervised network methods to develop the DL models, aiming to investigate and compare the model trained by using ROI information as input in SDLM and using whole CT images without manual lesion contours in WSDLM. Results showed that the SDLM with ROI information involved had better performance, which may be because the manual input of lesions annotation was necessary for DL model training to enhance signal-to-noise ratio with higher quality [26–28]. As reported, classification or prediction of DR-TB from CT images is considered a challenging task by the deep learning network [29]. The best accuracy rate for the classification of DR-TB was 0.516 in the *Tuberculosis Competition of ImageCLEF 2017* [29], and thereafter improved accuracy of 0.6–0.7 was reported [30,31]. The fusion model in the current study performed well in predicting success and DR-TB cases with AUC of 0.802 and 0.857 on internal and external datasets, which showed it has the potential to make TB treatment outcome predictions.

Unlike previously reported TB detection models focused on learning imaging features at a single timepoint [25,26,32], the current research involves longitudinal CT scans to track radiographic changes over time (Fig. 7). In the study, the models were significantly predictive of TB outcomes using early follow-up CT scans, which allows for assessing patient outcomes at an early stage of the treatment. This may be because longitudinal CT scans can provide more lesion characteristics than the static images obtained at a single timepoint and allow for extracting those subtle interval changes to make an accurate prediction. To the best of our knowledge, few studies have applied longitudinal CT images to develop a TB treatment outcome prediction model.

There are several limitations in the study. First, although an

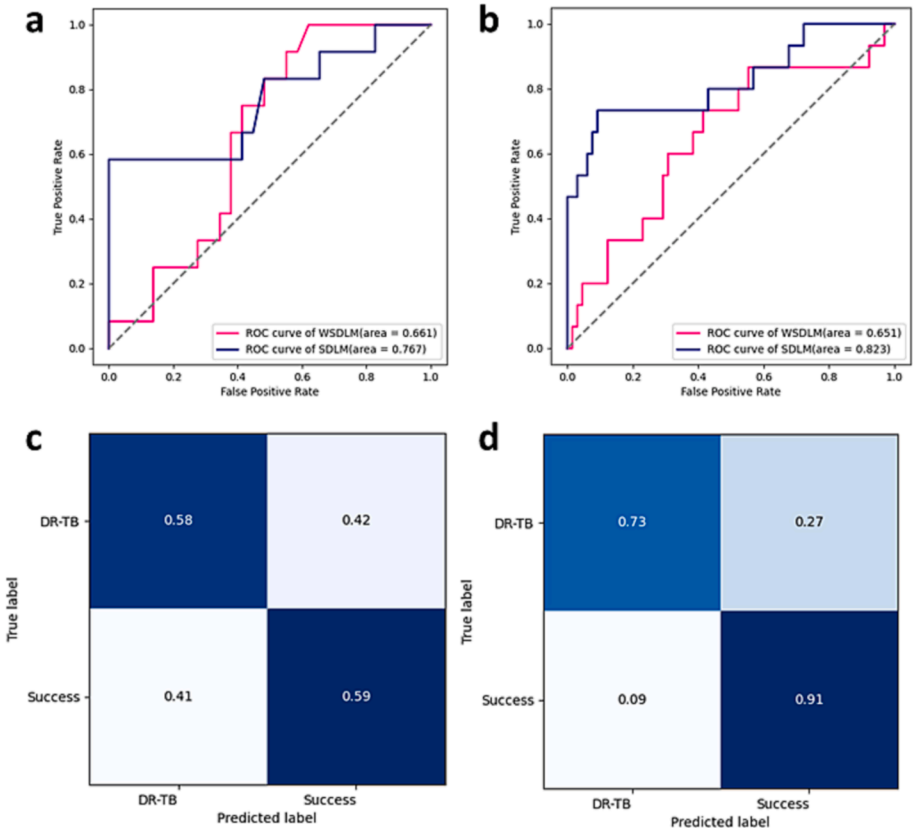

Fig. 5. Performance of the deep learning models in predicting TB treatment outcome. (a) ROC curves of two different deep learning models in the internal cohort. (b) ROC curves of two different deep learning models in the external cohort. (c) Normalized confusion matrix of the SDLM model in the external cohort. (d) Normalized confusion matrix of the SDLM model in the external cohort. TB, tuberculosis; ROC, receiver operating characteristic; SDLM, supervised deep learning model.

Table 2
Performance of the Bagging and SDLM to predict success and failure cases in internal and external cohorts.

Testing Dataset	AUC	*P* Value	Accuracy (95 % CI)	Sensitivity (95 % CI)	Specificity (95 % CI)	F1 Score (95 % CI)
Internal Cohort						
Bagging	0.789 (0.633–0.900)	0.002	0.805 (0.657–0.900)	0.667 (0.388–0.865)	0.862 (0.688–0.951)	0.667 (0.466–0.822)
SDLM	0.767 (0.609–0.885)	0.005	0.585 (0.434–0.723)	0.583 (0.319–0.807)	0.586 (0.407–0.745)	0.452 (0.292–0.622)
External Cohort						
Bagging	0.840 (0.633–0.900)	<0.001	0.800 (0.699–0.874)	0.733 (0.476–0.895)	0.815 (0.703–0.893)	0.579 (0.422–0.722)
SDLM	0.823 (0.609–0.885)	<0.001	0.875 (0.783–0.933)	0.733 (0.476–0.895)	0.908 (0.810–0.960)	0.688 (0.513–0.822)

Abbreviations: AUC, area under the curve; SDLM, supervised deep learning model.

independent testing was conducted, it only involved one external dataset. To objectively test the prediction model's generalizability and reliability, it is recommended to have it be validated on multi-center independent datasets. For the next step of our study, multiple external datasets will be involved and compared to test the prediction model. Second, although the fusion scheme performance has been improved, it may not be the optimal way to combine the radiomics feature method and the DL method. Thus, we should investigate and develop new fusion methods to fuse the different types of features in the next step.

5. Conclusions

This is the first study to propose and independently validate radiomics feature-based scheme and DL-based scheme to predict TB treatment outcomes using longitudinal CT scans. To further improve the prediction performance, we fused the prediction scores generated individually by the two schemes to build the new fusion model. Results showed that the fusion scheme yielded higher AUC values in distinguishing between DR-TB and success, which is clinically beneficial for patients at the early stage of the treatment.

Funding

This work was supported by the the Tianshan Innovation Team Program of Autonomous Region [grant number 2022D14007]; the National Natural Science Foundation of China [grant numbers82360359]; the Regional Collaborative Innovation Project of Xinjiang Uygur Autonomous Region (Shanghai Cooperation Organization Science and Technology Partnership Program & International Science and Technology Cooperation Programme) [Grant No. 2020E01012]; the National

Table 3
Performance of fusion models under different fusion strategies.

Method	Internal cohort		External cohort	
	AUC	95 % CI	AUC	95 % CI
Bagging	0.789	0.633–0.900	0.840	0.741–0.912
SDLM	0.767	0.609–0.885	0.823	0.721–0.899
Minimum	0.767	0.609–0.885	0.831	0.730–0.905
Maximum	0.802	0.648–0.910	0.857	0.761–0.926
0.1 × Bagging + 0.9 × SDLM	0.767	0.609–0.885	0.831	0.730–0.905
0.2 × Bagging + 0.8 × SDLM	0.796	0.641–0.906	0.835	0.735–0.909
0.3 × Bagging + 0.7 × SDLM	0.802	0.648–0.910	0.843	0.745–0.915
0.4 × Bagging + 0.6 × SDLM	0.790	0.635–0.901	0.850	0.753–0.920
0.5 × Bagging + 0.5 × SDLM	0.799	0.644–0.908	0.851	0.754–0.921
0.6 × Bagging + 0.4 × SDLM	0.793	0.638–0.903	0.856	0.760–0.925
0.7 × Bagging + 0.3 × SDLM	0.793	0.638–0.903	0.855	0.759–0.924
0.8 × Bagging + 0.2 × SDLM	0.790	0.635–0.901	0.857	0.761–0.926
0.9 × Bagging + 0.1 × SDLM	0.790	0.635–0.901	0.857	0.761–0.926

Abbreviations: AUC, area under the curve; SDLM, supervised deep learning model.

Key Research and Development Program of China [grant number 2019YFE0121400]; the Shenzhen Science and Technology Program [grant numbers KQTD2017033110081833, JSGG20201102162802008, JCYJ20220531093817040]; the Shenzhen Fundamental Research Program [grant number JCYJ20190813153413160]; and the Guangzhou Science and Technology Planning Project [grant number 2023A03J0536].

Ethical statement

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Institutional Review Board of The First People's Hospital of Kashi Prefecture ([2021]KDYYIRB(No.99)) and the patient consent was waived.

CRediT authorship contribution statement

Mayidili Nijiati: Writing – review & editing, Writing – original draft, Visualization, Conceptualization. **Lin Guo:** Conceptualization, Writing - original draft, Writing - review & editing. **Abudoukeyoumujiang Abulizi:** Conceptualization. **Shiyu Fan:** Resources. **Abulikemu**

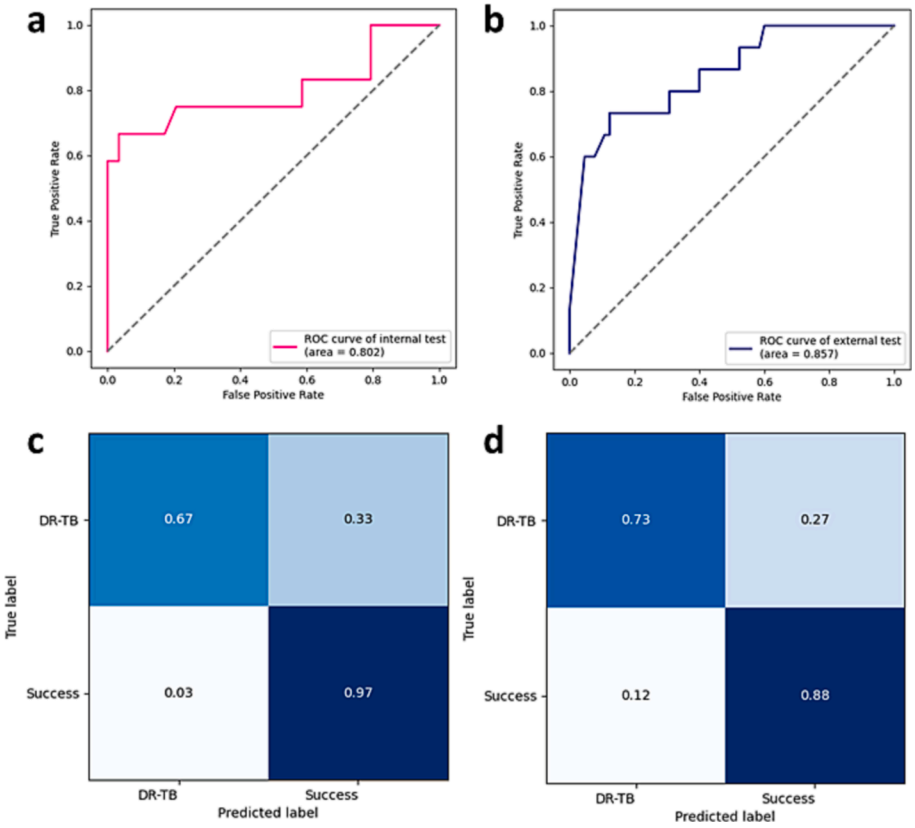

Fig. 6. Performance of the fusion model in predicting TB treatment outcomes of DR-TB cases and success cases. (a) ROC curves of the model in the internal cohort. (b) ROC curves of the model in the external cohort. (c) Normalized confusion matrix of the model in the internal cohort. (d) Normalized confusion matrix of the model in the external cohort. TB, tuberculosis; DT-TB, drug-resistant tuberculosis; ROC, receiver operating characteristic.

Table 4
Performance of the optimal fusion model to predict success and failure cases on internal and external cohorts.

Testing Dataset	AUC (95 % CI)	*P* value	Accuracy (95 % CI)	Sensitivity (95 % CI)	Specificity (95 % CI)	F1 Score (95 % CI)
Internal Cohort	0.802 (0.648–0.910)	0.002	0.878 (0.740–0.951)	0.667 (0.388–0.865)	0.966 (0.814–0.999)	0.762 (0.545–0.898)
External Cohort	0.857 (0.761–0.926)	<0.001	0.850 (0.754–0.914)	0.733 (0.476–0.895)	0.877 (0.773–0.939)	0.647 (0.479–0.786)

Abbreviations: AUC, area under the curve.

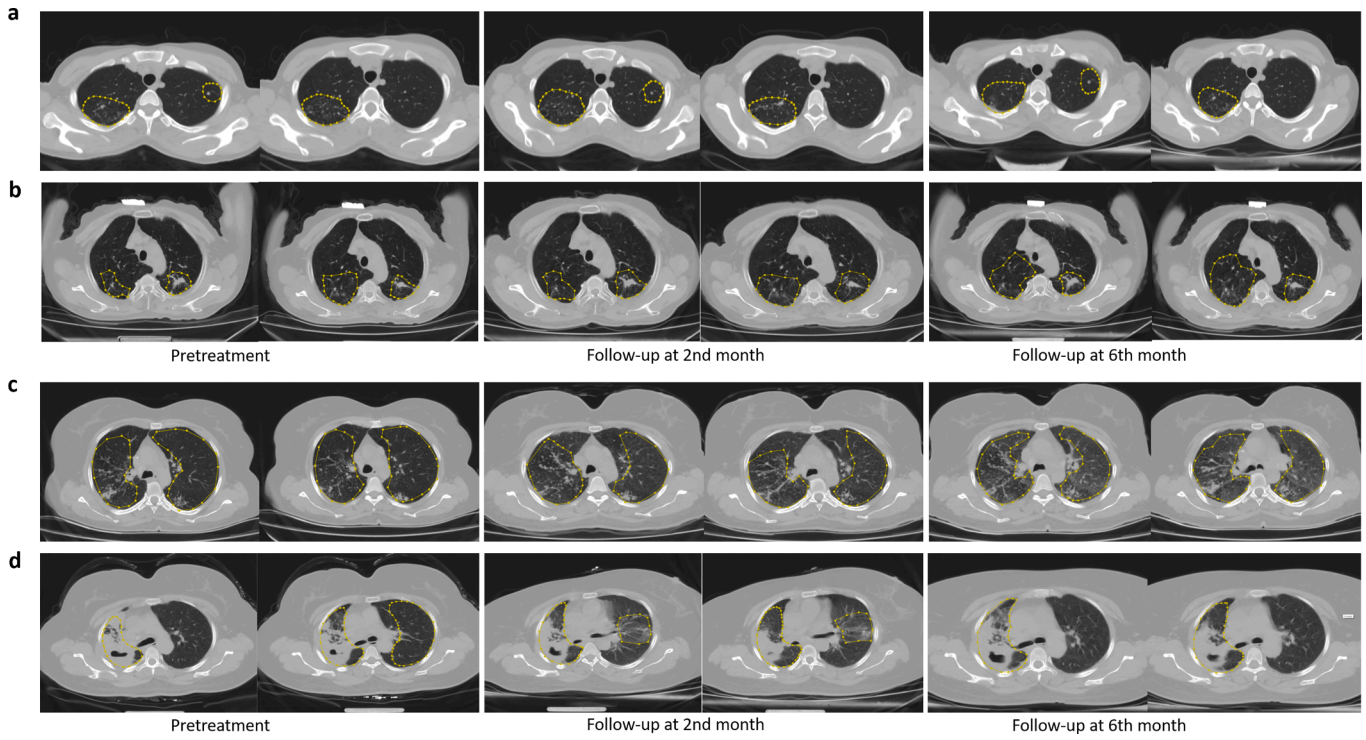

Fig. 7. Examples of radiographic changes of success cases and DR-TB cases arranging by time. (a) Pretreatment and 2 follow-up CT scans at the 2nd and 6th month of the treatment success case with ROI information. (b) 2 follow-up CT scans 2nd and 6th month of the DR-TB case with ROI information. DR-TB, drug-resistant tuberculosis; ROI, region of interest.

Wubuli: Resources. **Abudouresuli Tuersun:** Resources, Investigation. **Pahatijiang Nijiati:** Resources, Investigation. **Li Xia:** Formal analysis. **Kunlei Hong:** Methodology, Software. **Xiaoguang Zou:** Supervision.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All authors have completed the ICMJE uniform disclosure form. L.G. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed employee of Shenzhen Zhiying Medical Imaging. Other relationships: disclosed no relevant relationships. L.X. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed employee of Shenzhen Zhiying Medical Imaging. K.H. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: disclosed employee of Shenzhen Zhiying Medical Imaging. Other relationships: disclosed no relevant relationships. The other authors have no conflicts of interest to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrad.2023.111180.

References

[1] WHO. Global Tuberculosis Report 2022, 2022 (accessed 14 October 2021).
[2] G. Rosenfeld, A. Gabrielian, Q. Wang, J. Gu, D.E. Hurt, A. Long, A. Rosenthal, Radiologist observations of computed tomography (CT) images predict treatment outcome in TB Portals, a real-world database of tuberculosis (TB) cases, PLoS One 16 (2021) e0247906.
[3] S.-A.-O. Eshetie, A. Alebel, F. Wagnew, D. Geremew, A. Fasil, U. Sack, Current treatment of multidrug resistant tuberculosis in Ethiopia: an aggregated and individual patients' data analysis for outcome and effectiveness of the current regimens, BMC Infect Dis 18 (2018) 486.
[4] O.-A.-O. Hussain, K.-A.-O. Junejo, Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models, Inform Health Soc Care 44 (2019) 135–151.
[5] A. Gupta, V. Kumar, S. Natarajan, R. Singla, Adverse drug reactions & drug interactions in MDR-TB patients, Indian J Tuberc 67 (2020) 69–78.
[6] V.P. Keane, N. De Klerk, T. Krieng, G. Hammond, A.W. Musk, Risk factors for the development of non-response to first-line treatment for tuberculosis in southern Vietnam, Int. J. Epidemiol. 26 (1997) 1115–1120.
[7] C.M. Sauer, D.-A.-O. Sasson, K.-A.-O. Paik, N. McCague, L.A. Celi, I. Sánchez Fernández, B.M.W. Illigens, Feature selection and prediction of treatment failure in tuberculosis, PLoS One 13 (2018) e0207491.
[8] S.R.N. Kalhori, M. Nasehi, X.-J. Zeng, A Logistic Regression Model to Predict High Risk Patients to Fail in Tuberculosis Treatment Course Completion, IAENG Int J Appl Math 40 (2010) 102–107.
[9] T.M. Walker, T.A. Kohl, S.V. Omar, J. Hedge, E.C. Del Ojo, P. Bradley, Z. Iqbal, S. Feuerriegel, K.E. Niehaus, D.J. Wilson, D.A. Clifton, G. Kapatai, C.L.C. Ip, R. Bowden, F.A. Drobniewski, C. Allix-Béguec, C. Gaudin, J. Parkhill, R. Diel, P. Supply, D.W. Crook, E.G. Smith, A.S. Walker, N. Ismail, S. Niemann, T.E.A. Peto, Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study, Lancet Infect Dis 15 (2015) 1193–1202.
[10] K. Weyer, F. Mirzayev, G.B. Migliori, W. Van Gemert, L. Ambrosio, M. Zignol, K. Floyd, R. Centis, D.M. Cirillo, E. Tortoli, C. Gilpin, I.J. de Dieu, D. Falzon, M. Raviglione, Rapid molecular TB diagnosis: evidence, policy making and global implementation of Xpert MTB/RIF, Eur Respir J 42 (2013) 252.
[11] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R.H. Mak, H.J.W. L. Aerts, Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging, Clin Cancer Res 25 (2019) 3266–3275.
[12] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang, L. Ye, M. Gao, Z. Zhou, L. Li, J. Wang, Z. Yang, H. Cai, J. Xu, L. Yang, W. Cai, W. Xu, S. Wu, W. Zhang, S. Jiang, L. Zheng, X. Zhang, L. Wang, L. Lu, J. Li, H. Yin, W. Wang, O. Li, C. Zhang, L. Liang, T. Wu, R. Deng, K. Wei, Y. Zhou, T. Chen, J. Y. Lau, M. Fok, J. He, T. Lin, W. Li, G. Wang, Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography, Cell 181 (2020) 1423–1433.
[13] R.Y. Chen, L.E. Dodd, M. Lee, P. Paripati, D.A. Hammoud, J.M. Mountz, D. Jeon, N. Zia, H. Zahiri, M.T. Coleman, M.W. Carroll, J.D. Lee, Y.J. Jeong, P. Herscovitch, S. Lahouar, M. Tartakovsky, A. Rosenthal, S. Somaiyya, S. Lee, L.C. Goldfeder, Y. Cai, L.E. Via, S.K. Park, S.N. Cho, C.E. Barry 3rd, PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis, Sci. Transl. Med. 6 (2014) 265ra166.
[14] S.-A.-O. Malherbe, R.Y. Chen, P. Dupont, I. Kant, M. Kriel, A.G. Loxton, B. Smith, C. G.G. Beltran, S. van Zyl, S. McAnda, C. Abrahams, E. Maasdorp, A. Doruyter, L. E. Via, C.E. Barry 3rd, D. Alland, S.G. Richards, A. Ellman, T. Peppard, J. Belisle, G. Tromp, K. Ronacher, J.M. Warwick, J. Winter, G. Walzl, Quantitative 18F-FDG

PET-CT scan characteristics correlate with tuberculosis treatment response, EJNMMI Res 10 (2020) 8.
[15] X.-A.-O. Gao, Y. Qian, Prediction of multidrug-resistant TB from CT pulmonary images based on deep learning techniques, Mol Pharmacol 15 (2018) 4326–4335.
[16] Y. Li, B. Wang, L. Wen, H. Li, F. He, J. Wu, S. Gao, D. Hou, Machine learning and radiomics for the prediction of multidrug resistance in cavitary pulmonary tuberculosis: a multicentre study, Eur Radiol 19 (2022) 1–10.
[17] N.N. Linh, K. Viney, M. Gegia, D. Falzon, P. Glaziou, K. Floyd, H. Timimi, N. Ismail, M. Zignol, T. Kasaeva, F. Mirzayev, World Health Organization treatment outcome definitions for tuberculosis: 2021 update, Eur Respir J 58 (2021) 2100804.
[18] W. Zhou, G. Cheng, Z. Zhang, L. Zhu, S. Jaeger, F.Y.M. Lure, L. Guo, Deep learning-based pulmonary tuberculosis automated detection on chest radiography: large-scale independent testing, Quant Imag Med Surg 12 (2022) 2344–2355.
[19] F. Desissa, T. Workineh, T. Beyene, Risk factors for the occurrence of multidrug-resistant tuberculosis among patients undergoing multidrug-resistant tuberculosis treatment in East Shoa, Ethiopia. BMC Public Health 18 (2018) 422.
[20] Q. Huang, Y. Yin, S. Kuai, Y. Yan, J. Liu, Y. Zhang, Z. Shan, L. Gu, H. Pei, J. Wang, The value of initial cavitation to predict re-treatment with pulmonary tuberculosis, Eur J Med Res 21 (2016) 20.
[21] K.A. Mohammadzadeh, A. Ghayoomi, D. Maghsoudloo, Evaluation of factors associated with failure of tuberculosis treatment under DOTS in northern Islamic Republic of Iran, East Mediterr Health J 22 (2016) 87–94.
[22] B. Feng, X. Chen, Y. Chen, K. Liu, K. Li, X. Liu, N. Yao, Z. Li, R. Li, C. Zhang, J. Ji, W. Long, Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule, Eur J Radiol 128 (2020), 109022.
[23] E.N. Cui, T. Yu, S.J. Shang, X.Y. Wang, Y.L. Jin, Y. Dong, H. Zhao, Y.H. Luo, X. R. Jiang, Radiomics model for distinguishing tuberculosis and lung cancer on computed tomography scans, World J Clin Cases 8 (2020) 5203–5212.
[24] Y. Hu, X. Zhao, J. Zhang, J. Han, M. Dai, Value of 18F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis, Eur J Nucl Med Mol Imag 48 (2021) 231–240.
[25] L. Ma, Y. Wang, L. Guo, Y. Zhang, P. Wang, X. Pei, L. Qian, S. Jaeger, X. Ke, X. Yin, F.Y.M. Lure, Developing and verifying automatic detection of active pulmonary tuberculosis from multi-slice spiral CT images based on deep learning, J X-Ray Sci Technol 28 (2020) 939–951.
[26] C. Yan, L. Wang, J. Lin, J. Xu, T. Zhang, J. Qi, X. Li, W. Ni, G. Wu, J. Huang, Y. Xu, H.C. Woodruff, P. Lambin, A fully automatic artificial intelligence–based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis, Eur Radiol 32 (2022) 2188–2199.
[27] M. Nijiati, Z. Zhang, A. Abulizi, H. Miao, A. Tuluhong, S. Quan, L. Guo, T. Xu, X. Zou, Deep learning assistance for tuberculosis diagnosis with chest radiography in low-resource settings, J X-Ray Sci Technol 29 (2021) 785–796.
[28] E.-A.-O. Engle, A. Gabrielian, A. Long, D.-A.-O. Hurt, A. Rosenthal, Performance of Qure.ai automatic classifiers against a large annotated database of patients with diverse forms of tuberculosis, PLoS One 15 (2020) e0224445.
[29] Y.D. Cid, A. Kalinovsky, V. Liauchuk, V. Kovalev, H. Müller, Overview of ImageCLEF 2017 Tuberculosis Task – Predicting Tuberculosis Type and Drug Resistances, CLEF 2017 working Notes; Dublin, Ireland, 2017.
[30] S.-A.-O. Jaeger, O.H. Juarez-Espinosa, S.-A.-O. Candemir, M. Poostchi, F. Yang, L. Kim, M. Ding, L.R. Folio, S.-A.-O. Antani, A. Gabrielian, D. Hurt, A. Rosenthal, G. Thoma, Detecting drug-resistant tuberculosis in chest radiographs, Int J Comput Assist Radiol Surg 13 (2018) 1915–1925.
[31] M.H. Ali, D.M. Khan, K. Jamal, Z. Ahmad, S. Manzoor, Z. Khan, Prediction of multidrug-resistant tuberculosis using machine learning algorithms in SWAT, Pakistan, J Healthc Eng 2021 (2021) 2567080.
[32] S. Kulkarni, S. Jha, Artificial Intelligence, Radiology, and Tuberculosis: A Review, Acad Radiol 27 (2020) 71–75.