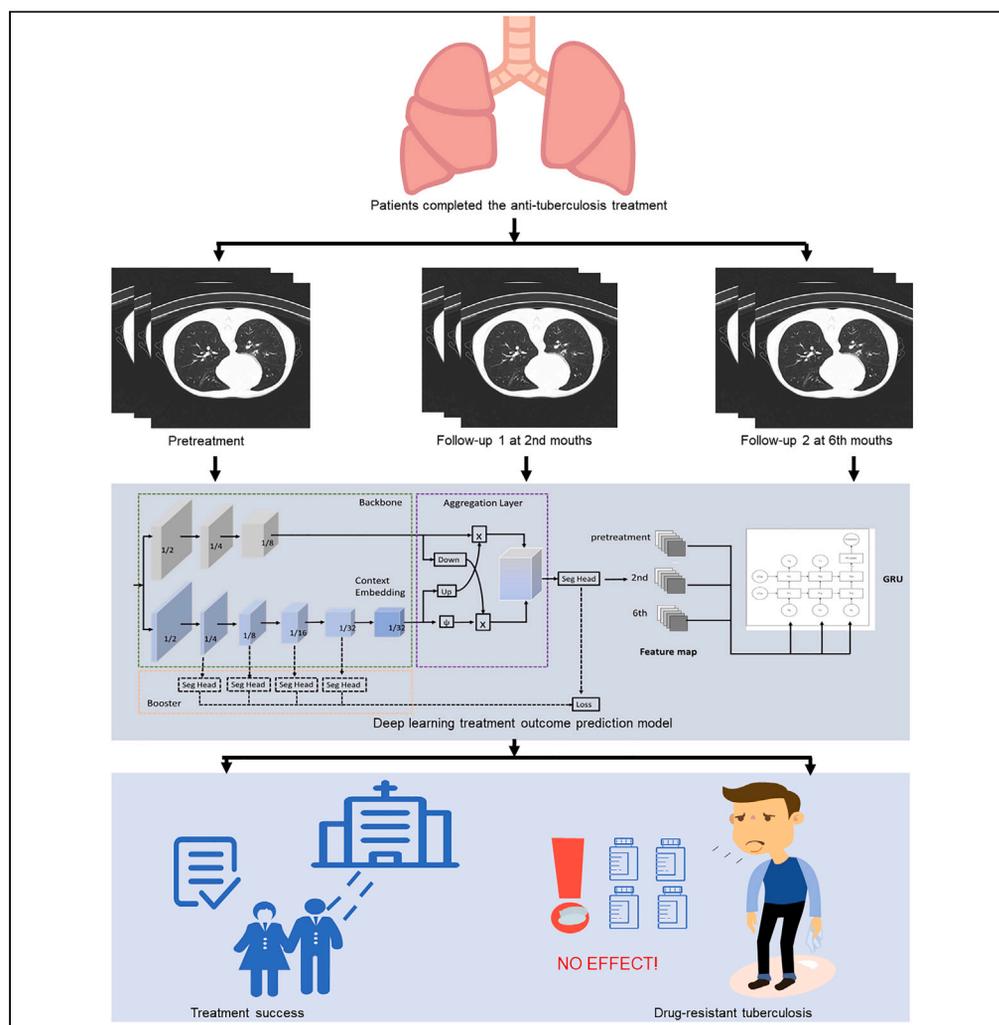


Article

# Deep learning on longitudinal CT scans: automated prediction of treatment outcomes in hospitalized tuberculosis patients



Mayidili Nijiati, Lin Guo, Abudouresuli Tuersun, ..., Li Xia, Kunlei Hong, Xiaoguang Zou

mydl0911@163.com (M.N.)  
zxgks@163.com (X.Z.)

**Highlights**

Drug-resistant outcome could be predicted in the early phases of therapy by imaging

Regular follow-up CT scans can aid in the treatment prediction of drug-resistant tuberculosis

Deep learning could successfully predict early tuberculosis treatment outcomes

Special attention should be given to the early phase of tuberculosis treatment

Nijiati et al., iScience 26, 108326  
November 17, 2023 © 2023 The Authors.  
<https://doi.org/10.1016/j.isci.2023.108326>



## Article

## Deep learning on longitudinal CT scans: automated prediction of treatment outcomes in hospitalized tuberculosis patients

Mayidili Nijati,<sup>1,4,\*</sup> Lin Guo,<sup>2,4</sup> Abudouresuli Tuersun,<sup>1,4</sup> Maihemitjiang Damola,<sup>1</sup> Abudoukeyoumujiang Abulizi,<sup>1</sup> Jiake Dong,<sup>1</sup> Li Xia,<sup>2</sup> Kunlei Hong,<sup>2</sup> and Xiaoguang Zou<sup>3,5,\*</sup>

## SUMMARY

**Three deep learning (DL)-based prediction models (PMs) using longitudinal CT images were developed to predict tuberculosis (TB) treatment outcomes. The internal dataset consists of 493 bacteriologically confirmed TB patients who completed the anti-tuberculosis treatment with three-time CT scans, including a pretreatment CT scan and two follow-up CT scans. PM1 was trained using only pretreatment CT scans, and PM2 and PM3 were developed by adding follow-up scans. An independent testing was performed on external dataset comprising 86 TB patients. The area under the curve for classifying success and drug-resistant (DR)-TB was improved on both internal (0.609 vs. 0.625 vs. 0.815) and external (0.627 vs. 0.705 vs. 0.735) dataset by adding follow-up scans. The accuracy and F1-score also showed an increasing tendency in the external test. Regular follow-up CT scans can aid in the treatment prediction, and special attention should be given to early intensive phase of treatment to identify high-risk DR-TB patients.**

## INTRODUCTION

Tuberculosis (TB) is a highly lethal infectious disease caused by the bacteria *Mycobacterium tuberculosis* that is extremely contagious. It is particularly frequent in resource-constrained situations, densely populated places, and locations with a high HIV prevalence. Despite the decline in TB incidence in previous years, the COVID-19 pandemic has reversed years of progress in providing essential TB services and reducing TB disease.<sup>1</sup> Therefore, it is crucial to take immediate action for TB treatment and control. Generally, TB patients who are enrolled on first-line treatment report a treatment success rate of 85%. However, drug-resistant TB (DR-TB) patients require longer and new treatment regimens due to their resistance to the first-line drugs.<sup>2,3</sup> Closer monitoring to identify early predictors of DR-TB is essential to help prevent such unsuccessful outcome and inform new treatment strategies. In most cases, the demographics and clinical data such as HIV, diabetes, alcohol use and adverse drug reactions are involved for DR-TB prediction,<sup>4,5</sup> but results vary depending on the setting and patient population, and different strongest predictive performances have been reported.<sup>6</sup> Whole-genome sequencing has also been used to characterize common and rare mutations that predict drug resistance.<sup>7</sup> However, this test is slow and expensive, which makes it less feasible for widespread use. Therefore, there is an urgent need for the development of rapid and affordable diagnostic tools for DR-TB prediction.

Chest CT imaging is a crucial tool in the management and monitoring of patients, and changes in CT images over time are essential indicators for predicting treatment outcomes.<sup>8–11</sup> Machine learning and radiomics have been reported to be applied on CT images to detect DR-TB patients, however, it focused on a single scan for the model input and manual labeling was needed each time the model was applied.<sup>12</sup> While artificial intelligence (AI) techniques have shown promise in predicting treatment outcomes for COVID-19 and lung cancer based on CT images,<sup>8,13</sup> very few studies have applied deep learning (DL) networks to predict TB treatment outcomes using serial CT images.

Given that poor treatment outcomes are common in the early phases of therapy,<sup>14</sup> we developed three DL models using serial CT scans (including pretreatment and posttreatment CT scans at the second and the sixth months follow-up) to predict the TB treatment outcomes of success and DR-TB cases in the study. The first model used only pretreatment CT scans, while the other two included pretreatment scans and one or two follow-up scans to explore the role of serial CT images and how early DR-TB outcomes could be predicted. The objectives of our research focus on the identification of TB patients with the potential to progress into DR-TB, and determining the feasibility of predicting early DR-TB treatment outcomes using CT imaging data. We tested the proposed models in an internal set and validated them independently on an external set. This study is exceptionally beneficial for patients who are at risk of developing DR-TB and could lead to improved patient outcomes.

<sup>1</sup>Department of Radiology, The First People's Hospital of Kashi Prefecture, Kashi, China

<sup>2</sup>Shenzhen Zhiying Medical Imaging, Shenzhen, China

<sup>3</sup>Clinical Medical Research Center, The First People's Hospital of Kashi Prefecture, Kashi, China

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead contact

\*Correspondence: mydl0911@163.com (M.N.), zxgks@163.com (X.Z.)

<https://doi.org/10.1016/j.isci.2023.108326>



**Table 1. Patient demographics and treatment outcomes of 2 different datasets**

Demographics	Site A (n = 493)	Site B (n = 86)	$T/\chi^2$	p Value
Age (yr, mean $\pm$ sd.)	58.66 $\pm$ 18.66	58.85 $\pm$ 17.80	0.088	0.93
Sex (n male)	267 (51.2%)	35(40.7%)	5.317	0.021
<b>Outcomes</b>				
Success (n (%))	410 (83.2%)	56 (65.1%)	15.185	<0.001 <sup>a</sup>
Cured (n (%))	260 (52.7%)	43 (50.0%)	18.423	<0.001 <sup>b</sup>
Treatment completed (n (%))	150 (30.5%)	13 (15.1%)	NA	NA
Failed (n (%))			NA	NA
Transferred to drug-resistant therapy (n (%))	83 (16.8%)	30 (34.9%)	NA	NA

NA, not applicable.

<sup>a</sup>Comparison of outcomes of success and failed.

<sup>b</sup>Comparison of outcomes of cured, treatment completed, transferred to drug-resistant.

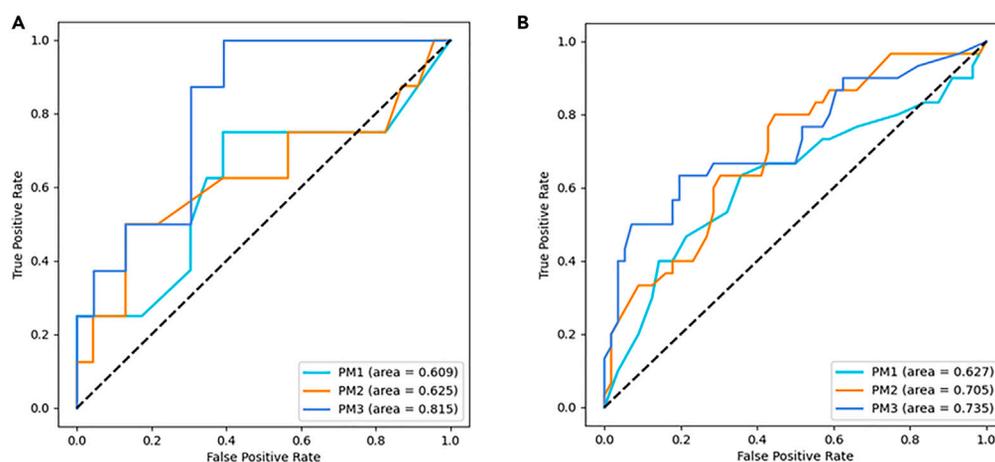
## RESULTS

### Clinical characteristics

The main characteristics of patients included in both the internal and external datasets are shown in Table 1. These two cohorts displayed inherent differences. The internal dataset from Site A was utilized to develop and evaluate the performance of three deep learning-based prediction models, while the external dataset from Site B was used as an independent validation set to test the proposed models' generalization ability. The internal dataset comprised a total of 493 patients from Site A, with 51.2% of them being male, and a median age of 58.66  $\pm$  18.66 years. On the other hand, the external dataset contained 86 patients from Site B, with 40.7% being male, and a median age of 58.85  $\pm$  17.80 years. Although there was no significant difference in patient age between the two datasets ( $p > 0.05$ , Table 1), a significant difference was observed in patient gender ( $p < 0.05$ , Table 1).

### Performance evaluation of PM1, PM2, and PM3 in internal testing

The objective of our study was to identify patients who would fail TB treatment and develop DR-TB by training deep learning-based prediction models using longitudinal CT scans from Site A. Our findings, presented in Figure 1, demonstrate that PM1, which relied solely on pre-treatment scans, showed the lowest performance in predicting DR-TB (AUC = 0.609,  $p = 0.423$ ). However, the addition of follow-up scans significantly improved the performance of PM2 (AUC = 0.625,  $p = 0.335$ ) and PM3 (AUC = 0.815,  $P < 0.001$ ). Over the course of the first two months, the AUC increased by 0.016 (PM1 vs. PM2), followed by a further increase of 0.19 in the subsequent four months (PM2 vs. PM3). Optimal thresholds of 0.29, 0.30, and 0.25 were applied to calculate evaluation metrics of accuracy, sensitivity, specificity, and F1 score for PM1, PM2, and PM3, which were 0.645, 0.750, 0.609, and 0.552 for PM1, 0.581, 0.625, 0.565, and 0.435 for PM2, and 0.742, 0.875, 0.696, and 0.636 for PM3 (Table 2).



**Figure 1. The comparison of the ROC curves of PM1, PM2, and PM3 on different datasets**

(A) Internal test.

(B) External test.

**Table 2. Performance of three different prediction models on internal and external datasets**

Test	AUC (95% CI)	p value	Accuracy		Sensitivity		Specificity		F1 Score	
			(95% CI)	p value	(95% CI)	p value	(95% CI)	p value	(95% CI)	p value
Internal test										
PM1	0.609 (0.418–0.778)	0.423	0.645 (0.469–0.790)		0.750 (0.401–0.937)		0.609 (0.407–0.779)		0.522 (0.330–0.708)	
PM2	0.625 (0.434–0.791)	0.355	0.581 (0.407–0.736)	0.602 <sup>a</sup>	0.625 (0.304–0.865)	0.590 <sup>a</sup>	0.565 (0.368–0.744)	0.765 <sup>a</sup>	0.435 (0.256–0.632)	0.555 <sup>a</sup>
PM3	0.815 (0.635–0.931)	< 0.001	0.742 (0.565–0.865)	0.180 <sup>b</sup>	0.875 (0.508–0.999)	0.248 <sup>b</sup>	0.696 <sup>b</sup> (0.489–0.846)	0.359 <sup>b</sup>	0.636 (0.429–0.804)	0.175 <sup>b</sup>
External test										
PM1	0.627 (0.516–0.729)	0.063	0.523 (0.419–0.625)		0.733 (0.555–0.860)		0.411 (0.292–0.541)		0.489 (0.388–0.591)	
PM2	0.705 (0.597–0.799)	0.001	0.651 (0.546–0.744)	0.088 <sup>a</sup>	0.633 (0.455–0.782)	0.405 <sup>a</sup>	0.661 (0.530–0.771)	0.008 <sup>a</sup>	0.559 (0.441–0.671)	0.384 <sup>a</sup>
PM3	0.735 (0.629–0.825)	< 0.001	0.686 (0.582–0.775)	0.627 <sup>b</sup>	0.667 (0.487–0.809)	0.787 <sup>b</sup>	0.696 (0.566–0.802)	0.686 <sup>b</sup>	0.597 (0.477–0.706)	0.653 <sup>b</sup>

<sup>a</sup>Comparison of PM1 and PM2.

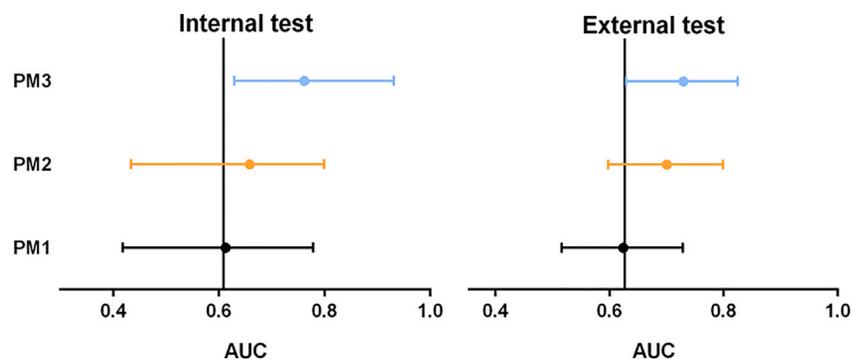
<sup>b</sup>Comparison of PM2 and PM3.

### Performance evaluation of PM1, PM2, and PM3 in external testing

In order to evaluate the generalizability of our prediction models of the PM1, PM2, and PM3, the independent dataset Site B was involved as the external testing dataset. Consistent with the results of the internal test, we observed an improvement in the models' performance with each additional CT scan used for training (Figure 1). The comprehensive metrics of the AUC, accuracy, and F1 score metrics showed an increase of 0.078 and 0.03, 0.128 and 0.034, 0.07 and 0.038 for PM1 vs. PM2 and PM2 vs. PM3 comparisons, respectively.

The results depicted in Figure 2 demonstrate that the performance of the prediction models improved more prominently during the first two-month intensive phase of tuberculosis treatment as compared to the subsequent four-month continuation phase. This finding is noteworthy as it suggests that close monitoring and intervention during the initial intensive phase of treatment may have a more significant impact on patient outcomes, particularly in identifying patients who are at a higher risk of developing drug-resistant tuberculosis. It is important to note, however, that this trend was not consistently observed in the internal test dataset. This discrepancy may be attributed to several factors, including the smaller sample size in the internal set (31 vs. 86) and the differences in population or characteristics between the two datasets. Notably, the distribution of success and DR-TB outcomes in the internal dataset was significantly different from that of the external dataset (Tables 1 and 3).

In the external test, the three thresholds were set the same as they were in the internal set, and the sensitivity and specificity were 0.733 and 0.411 for the PM1, 0.633 and 0.661 for the PM2, and 0.667 and 0.696 for the PM1. Additionally, we presented representative cases of a false-positive and a false-negative result detected by both PM2 and PM3 in Figure 3. Our study suggests that follow-up scans may aid in identifying unsuccessful outcomes of DR-TB and informs new treatment strategies to improve patient outcomes. Furthermore, the findings indicate that close attention should be paid to the early anti-tuberculosis therapy, as the imaging from the first two months or even earlier in the treatment process has shown a trend in predicting the treatment outcome. This emphasizes the importance of early detection and intervention to prevent the progression of DR-TB and improve patient outcomes.



**Figure 2. The comparison of the AUC of three different prediction models on internal and external datasets**

**Table 3. WHO definition of treatment outcomes for TB patients**

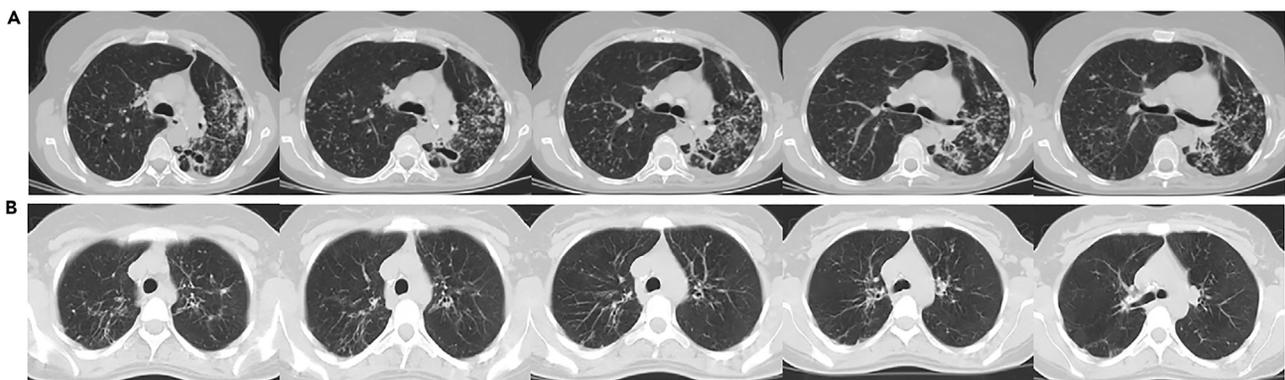
Outcome	Definition
Cured	TB patients with bacteriologically confirmed TB at the beginning of treatment which was smear-or-culture-negative in the last month of treatment and on at least one previous occasion
Treatment completed	TB patients who completed treatment without evidence of failure but with no record to show that sputum smear or cultures results in the last month of treatment and on at least one previous occasion were negative
Treatment success	Composite of cured and treatment completed
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 and extensively drug-resistant TB

TB, tuberculosis; DR-TB, drug-resistant TB.

## DISCUSSION

In this study, we aimed to develop three DL models to predict TB treatment outcomes using longitudinal CT images, and both internal and external testing was conducted to evaluate their performance, and the main advancements of the research are that we found that DR-TB outcome could be predicted in the early phases of therapy by imaging, and close monitoring and intervention during the initial intensive phase of treatment may have a more significant impact on patient outcomes.

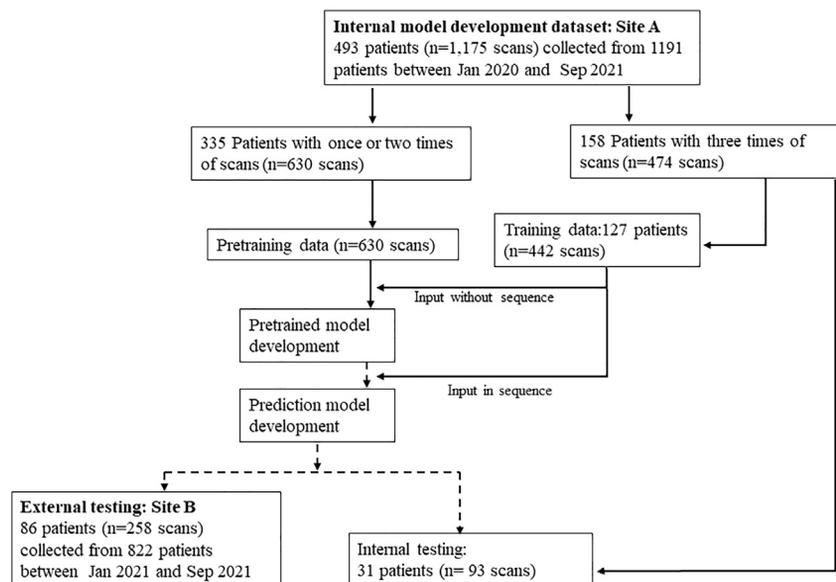
DR-TB means the bacterium *Mycobacterium TB* that is resistant to routinely used anti-TB medications. This is one of the major obstacles in the global fight against tuberculosis because it may complicate therapy and raise the probability of poor results. Several research studies have focused on developing tools and methods to predict drug resistance in TB from whole genome sequencing,<sup>15</sup> phylogenomic,<sup>16</sup> nomogram<sup>17</sup> to clinical implications.<sup>18</sup> Conventionally, in the studies related to tuberculosis treatment, clinical and demographic features were used to predict treatment outcomes, where the number of previous treatments, lack of a job, and alcohol consumption were identified as potential risk factors for the occurrence of DR-TB,<sup>19</sup> and a model based on these features was reported to achieve an AUC of 0.74.<sup>20</sup> Clinical and demographic information, while useful, do not incorporate phenotypic changes, and moreover, various reported studies have identified different factors as the most effective predictors,<sup>21</sup> with the AUC ranging from 0.6 to 0.8.<sup>22–26</sup> This lack of consistency in the results among studies, where different factors have been identified as the strongest predictors, poses a challenge for identifying a universal factor that can be practically applied in clinical settings.<sup>20,27,28</sup> Instead of using the clinical and demographic features, we proposed DL models using CT images, and the PM3 with the best performance in this study achieved a high performance in predicting success and treatment failure cases, achieving AUC values of 0.815 and 0.735 on internal and external datasets, respectively. We observed that follow-up CT scans can provide more lesion characteristics and subtle interval changes, which may help to accurately predict TB treatment outcomes. Therefore, compared to the poorest performance of the PM1 model, the addition of follow-up scans resulted in an improved prediction performance in the PM2 and PM3, which may help prevent patients from evolving into poor outcomes during treatment. Our findings demonstrate the potential of using DL models in predicting TB treatment outcomes based on longitudinal CT scans and may have important implications for improving patient care in clinical settings.



**Figure 3. Examples of a false-positive case and a false-negative case of the proposed model for predicting DR-TB from CT images**

(A) A false-positive case: the proposed model misdiagnosed a success case as DR-TB.

(B) A false-negative case: the proposed model misdiagnosed a DR-TB case as success case. DR-TB, drug-resistant tuberculosis.



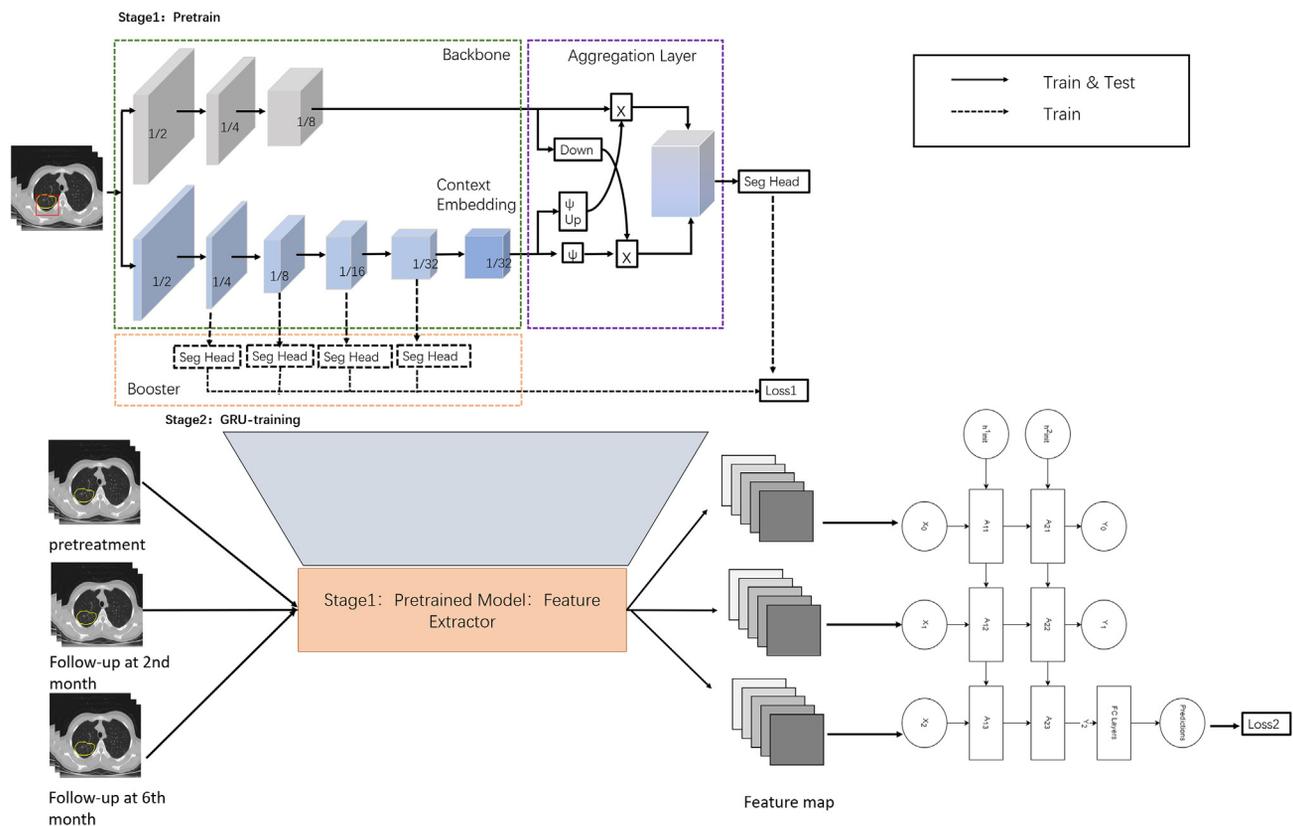
**Figure 4. Flowchart for study dataset**

A total of 2,013 patients were involved in the evaluation, and 579 patients were finally included, of which 493 patients were used as internal dataset and 86 patients were for the external validation. The solid line indicates data flowchart and the dotted line indicates model development flowchart.

Radiological images have been extensively investigated in DL models in TB, and previous studies have predominantly focused on TB detection based on chest X-rays or CT images, due to the availability of large public datasets that can be used to train automatic diagnosis systems. However, most of the public datasets often contain only static images collected at a single time point, so the models developed by previous studies mainly focused on TB diagnosis,<sup>29–31</sup> resulting in limiting the ability of the models to predict treatment outcomes. Effective diagnosis and prognostic evaluation are critical components of clinical workflows for TB control. Generally, TB can be cured within 6 months if the patient follows the treatment process.<sup>32</sup> However, treatment processes can vary from patient to patient and drug resistance can develop. To address these limitations, this current research utilizes longitudinal CT scans (collected in-house) to track radiographic changes over time, which enables the assessment of patient outcomes at an early stage of therapy.

In addition to the observation of model improvement with the use of follow-up scans, we noted that the greatest improvement occurred during the first two-month intensive phase of anti-tuberculosis treatment, as evidenced by the comprehensive evaluation metrics of AUC, accuracy, and F1 score in the external test. This finding suggests that monitoring the treatment process during the initial intensive phase of treatment may be particularly important for TB patients. Therefore, it is crucial to pay special attention to the anti-tuberculosis therapy effect during this period of time. Though the internal test did not reveal such a trend, which might be caused by the smaller number of patients in the internal set (31 vs. 86 in the external set) and differences in population characteristics may account for this discrepancy, as a significant change in the distribution of success and DR-TB outcomes was observed between the two sites (Figure 4). Our finding may provide radiological evidence for the treatment monitoring process with official guidelines that have indicated early identification of treatment failure is crucial for successful treatment of TB. The WHO recommends frequent monitoring of patient during the intensive phase of treatment to ensure that they are responding appropriately to therapy and to detect any potential treatment failures.<sup>33</sup> This may include regular sputum microscopy, culture, and drug susceptibility testing, as well as chest radiography or computed tomography if available. Additionally, the Centers for Disease Control and Prevention (CDC) also recommend close monitoring of TB patients during the initial phase of treatment. They state that patients should have clinical evaluations at least monthly to identify possible adverse effects of the anti-TB medications and to assess adherence, and it is critical to obtain a sputum specimen at the end of the intensive phase (2 months) to determine if the continuation phase should be extended.<sup>34</sup>

In the current study, a relatively small serial CT dataset was involved for the model construction, and to achieve a desirable performance on limited data, we applied BiSeNetV2-3D to make the most use of imaging information in a fast and accurate way (Figure 5). Only limited studies using BiSeNetV2 for COVID-19 segmentation,<sup>35</sup> and this is the first try to utilize BiSeNetV2 to identify the lesion location and extract features. Classification or prediction of DR-TB from CT images is considered a difficult and challenging task by the deep learning network.<sup>36</sup> The best accuracy rate for the classification of DR-TB was 0.516 in the *Tuberculosis Competition of ImageCLEF 2017*,<sup>36</sup> and thereafter improved accuracy of 0.6–0.7 was reported.<sup>37,38</sup> Our model achieved a comparable accuracy of 0.651 and 0.686 on the external dataset for PM2 and PM3, respectively. Recently, a higher accuracy over 0.7 (0.720–0.767) has been reported using both machine learning (ML) and radiomics, which suggested it might be a good way to incorporate the DL/ML and radiomics to develop a more accurate model identifying DR-TB.<sup>39</sup>



**Figure 5. The structure of BiSeNetV2-3D and the development of PMs**

Stage 1: The dual-channel backbone has a detail branch (gray dataset) and a semantic branch (blue dataset). The last stage of semantic branching is the output of context embedded block. Meanwhile, the number in the cube is the ratio of the feature mapping size to the input resolution. In the part of polymerization layer, we adopted bilateral polymerization layer. The sampling operation shows that Up indicates up-sampling operation and Sigmoid function was used, and x means element-wise output. In addition, some auxiliary segmentation heads were designed to improve the segmentation performance. Stage 2: The serial CT scans were input into pretrained model in sequence for feature extraction and those feature maps were further input into GRU for classification.

In this study, we have pioneered the development of DL models for the prediction of TB treatment outcomes based on longitudinal CT scans. Our findings suggest that regular follow-up CT scans can aid in the prediction of DR-TB, and that special attention should be given to the early intensive phase of treatment to identify high-risk DR-TB patients, which could have significant clinical benefits.

### Limitations of the study

There are several limitations in the study. First, special attention should be paid to the generalizability of the proposed models, especially when patient population and characteristics differ significantly between datasets. In the study, despite the consistent improvement in AUC observed for both internal and external testing, there was a difference in the improvement rate in the end. Therefore, further research employing multiple datasets to develop and validate a more reliable model is necessary. Secondly, although an independent test was conducted, it only involved a single external dataset, and there is a need for validation on multiple independent datasets from different centers.<sup>40</sup> Hence, for the follow-up works, further validation and exploration in larger datasets are needed, and we plan to incorporate various large-scale external datasets and compare them to evaluate the prediction system thoroughly.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ethical statement
- METHOD DETAILS
  - Patient cohorts
  - CT image acquisition and image preprocessing
  - Development of the PM1, PM2 and PM3
- QUANTIFICATION AND STATISTICAL ANALYSIS

## ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China [grant number 2022ZD0160705]; Tianshan Innovation Team Program of Autonomous Region [grant number 2022D14007]; Kashi Regional Science and Technology Plan Project [grant number KS2022017], the State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia [grant number SKL-HIDCA-2021-JH6]; the National Key Research and Development Program of China [grant number 2019YFE0121400]; the Shenzhen Science and Technology Program [grant numbers KQTD2017033110081833, JSYG20201102162802008, JCYJ20220531093817040]; the Shenzhen Fundamental Research Program [grant number JCYJ20190813153413160] and the Guangzhou Science and Technology Planning Project [grant number 2023A03J0536].

## AUTHOR CONTRIBUTIONS

M.N.: Conceptualization, writing – original draft, resources, and writing – reviewing and editing; L.G.: Conceptualization, writing – original draft, and methodology; A.T.: Data curation and investigation; M.D.: Data curation and formal analysis; A.A.: Data curation and formal analysis; J.D.: Investigation and data curation; L.X.: Visualization, investigation, and formal analysis; K.H.: Software and validation; X.Z.: Supervision, writing – reviewing and editing, and funding acquisition.

## DECLARATION OF INTERESTS

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.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: May 14, 2023

Revised: August 17, 2023

Accepted: October 20, 2023

Published: October 23, 2023

## REFERENCES

1. Organization, W.H. (2022). Global Tuberculosis Report 2022 (World Health Organization Press).
2. Eshetie, S., Alebel, A., Wagnew, F., Geremew, D., Fasil, A., and Sack, U. (2018). 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, 486.
3. Hussain, O.A., and Junejo, K.N. (2019). Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models. *Inf. Health Soc. Care* 44, 135–151.
4. Abdelbary, B.E., Garcia-Viveros, M., Ramirez-Oropesa, H., Rahbar, M.H., and Restrepo, B.I. (2017). Predicting treatment failure, death and drug resistance using a computed risk score among newly diagnosed TB patients in Tamaulipas, Mexico. *Epidemiol. Infect.* 145, 3020–3034.
5. Baluku, J.B., Nakazibwe, B., Naloka, J., Nabwana, M., Mwanja, S., Mulwana, R., Sempiira, M., Nassozi, S., Babirye, F., Namugenyi, C., et al. (2021). Treatment outcomes of drug resistant tuberculosis patients with multiple poor prognostic indicators in Uganda: A countrywide 5-year retrospective study. *J. Clin. Tuberc. Other Mycobact. Dis.* 23, 100221.
6. Peetluk, L.S., Ridolfi, F.M., Rebeiro, P.F., Liu, D., Rolla, V.C., and Sterling, T.R. (2021). Systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults. *BMJ Open* 11, e044687.
7. Walker, T.M., Kohl, T.A., Omar, S.V., Hedge, J., Del Ojo Elias, C., Bradley, P., Iqbal, Z., Feuerriegel, S., Niehaus, K.E., Wilson, D.J., et al. (2015). Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect. Dis.* 15, 1193–1202.
8. Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., Mak, R.H., and Aerts, H.J.W.L. (2019). Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clin. Cancer Res.* 25, 3266–3275.
9. Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., Zha, Y., Liang, W., Wang, C., Wang, K., et al. (2020). Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* 182, 1360–1433.
10. Chen, R.Y., Dodd, L.E., Lee, M., Paripati, P., Hammoud, D.A., Mountz, J.M., Jeon, D., Zia, N., Zahiri, H., Coleman, M.T., et al. (2014). PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis. *Sci. Transl. Med.* 6, 265ra166.
11. Malherbe, S.T., Chen, R.Y., Dupont, P., Kant, I., Kriel, M., Loxton, A.G., Smith, B., Beltran, C.G.G., van Zyl, S., McAnda, S., et al. (2020).

- Quantitative 18F-FDG PET-CT scan characteristics correlate with tuberculosis treatment response. *EJNMMI Res.* 10, 8.
12. Gao, X.W., and Qian, Y. (2018). Prediction of multidrug-resistant TB from CT pulmonary images based on deep learning techniques. *Mol. Pharm.* 15, 4326–4335.
  13. Feng, Z., Shen, H., Gao, K., Su, J., Yao, S., Liu, Q., Yan, Z., Duan, J., Yi, D., Zhao, H., et al. (2021). Machine learning based on clinical characteristics and chest CT quantitative measurements for prediction of adverse clinical outcomes in hospitalized patients with COVID-19. *Eur. Radiol.* 31, 7925–7935.
  14. Ketema, D.B., Muchie, K.F., and Andargie, A.A. (2019). Time to poor treatment outcome and its predictors among drug-resistant tuberculosis patients on second-line anti-tuberculosis treatment in Amhara region, Ethiopia: retrospective cohort study. *BMC Publ. Health* 19, 1481.
  15. Liu, D., Huang, F., Zhang, G., He, W., Ou, X., He, P., Zhao, B., Zhu, B., Liu, F., Li, Z., et al. (2022). Whole-genome sequencing for surveillance of tuberculosis drug resistance and determination of resistance level in China. *Clin. Microbiol. Infect.* 28, 731.e9–731.e15.
  16. Namburete, E.I., Dippenaar, A., Conceição, E.C., Feliciano, C., Nascimento, M.M.P.d., Peronni, K.C., Silva, W.A., Jr., Ferro, J.J., Harrison, L.H., Warren, R.M., and Bollela, V.R. (2020). Phylogenomic assessment of drug-resistant *Mycobacterium tuberculosis* strains from Beira, Mozambique. *Tuberculosis* 121, 101905.
  17. Wang, S., and Tu, J. (2020). Nomogram to predict multidrug-resistant tuberculosis. *Ann. Clin. Microbiol. Antimicrob.* 19, 27.
  18. Mnyambwa, N.P., Kim, D.J., Ngadaya, E.S., Kazwala, R., Petrucka, P., and Mfinanga, S.G. (2017). Clinical implication of novel drug resistance-conferring mutations in resistant tuberculosis. *Eur. J. Clin. Microbiol. Infect. Dis.* 36, 2021–2028.
  19. Desissa, F., Workneh, T., and Beyene, T. (2018). Risk factors for the occurrence of multidrug-resistant tuberculosis among patients undergoing multidrug-resistant tuberculosis treatment in East Shoa, Ethiopia. *BMC Publ. Health* 18, 422.
  20. Sauer, C.M., Sasson, D., Paik, K.E., McCague, N., Celi, L.A., Sánchez Fernández, I., and Illigens, B.M.W. (2018). Feature selection and prediction of treatment failure in tuberculosis. *PLoS One* 13, e0207491.
  21. Mnyambwa, N.P., Kim, D.J., Ngadaya, E.S., Kazwala, R., Petrucka, P., and Mfinanga, S.G. (2017). Clinical implication of novel drug resistance-conferring mutations in resistant tuberculosis. *Eur. J. Clin. Microbiol. Infect. Dis.* 36, 2021–2028. <https://doi.org/10.1007/s10096-017-3027-3>.
  22. Belilovsky, E.M., Borisov, S.E., Cook, E.F., Shaykevich, S., Jakubowiak, W.M., and Kourbatova, E.V. (2010). Treatment interruptions among patients with tuberculosis in Russian TB hospitals. *Int. J. Infect. Dis.* 14, e698–e703.
  23. Chang, K.C., Leung, C.C., and Tam, C.M. (2004). Risk factors for defaulting from anti-tuberculosis treatment under directly observed treatment in Hong Kong. *Int. J. Tubercul. Lung Dis.* 8, 1492–1498.
  24. Baussano, I., Pivetta, E., Vizzini, L., Abbona, F., and Bugiani, M. (2008). Predicting tuberculosis treatment outcome in a low-incidence area. *Int. J. Tubercul. Lung Dis.* 12, 1441–1448.
  25. Costa-Veiga, A., Briz, T., and Nunes, C. (2018). Unsuccessful treatment in pulmonary tuberculosis: factors and a consequent predictive model. *Eur. J. Publ. Health* 28, 352–358.
  26. Mburu, J.W., Kingwara, L., Ester, M., and Andrew, N. (2018). Use of classification and regression tree (CART), to identify hemoglobin A1C (HbA1C) cut-off thresholds predictive of poor tuberculosis treatment outcomes and associated risk factors. *J. Clin. Tuberc. Other Mycobact. Dis.* 11, 10–16.
  27. Huang, Q., Yin, Y., Kuai, S., Yan, Y., Liu, J., Zhang, Y., Shan, Z., Gu, L., Pei, H., and Wang, J. (2016). The value of initial cavitation to predict re-treatment with pulmonary tuberculosis. *Eur. J. Med. Res.* 21, 20.
  28. Mohammadzadeh, K.A., Ghayoomi, A., and Maghsoudloo, D. (2016). Evaluation of factors associated with failure of tuberculosis treatment under DOTS in northern Islamic Republic of Iran. *East. Mediterr. Health J.* 22, 87–94.
  29. Kulkarni, S., and Jha, S. (2020). Artificial Intelligence, Radiology, and Tuberculosis: A Review. *Acad. Radiol.* 27, 71–75.
  30. Yan, C., Wang, L., Lin, J., Xu, J., Zhang, T., Qi, J., Li, X., Ni, W., Wu, G., Huang, J., et al. (2022). A fully automatic artificial intelligence-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. *Eur. Radiol.* 32, 2188–2199.
  31. Ma, L., Wang, Y., Guo, L., Zhang, Y., Wang, P., Pei, X., Qian, L., Jaeger, S., Ke, X., Yin, X., and Lure, F.Y.M. (2020). 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, 939–951.
  32. Tiwari, A., and Maji, S. (2019). Machine Learning Techniques for Tuberculosis Prediction. In *International Conference on Advances in Engineering Science Management & Technology (ICAESMT)*.
  33. Organization, W.H. (2020). WHO Consolidated Guidelines on Tuberculosis: Module 4: Treatment: Drug-Resistant Tuberculosis Treatment.
  34. Prevention, C.f.D.C.a. (2021). Treatment of Tuberculosis (TB).
  35. Ma, S., Li, X., Tang, J., and Guo, F. (2022). EAA-Net: Rethinking the Autoencoder Architecture with Intra-class Features for Medical Image Segmentation. Preprint at ArXiv 1. <https://doi.org/10.48550/arXiv.2208.09197>.
  36. Cid, Y.D., Kalinovsky, A., Liauchuk, V., Kovalev, V., and Müller, H. (2017). Overview of ImageCLEF 2017 Tuberculosis Task – Predicting Tuberculosis Type and Drug Resistances. CLEF 2017 Working Notes.
  37. Jaeger, S., Juarez-Espinosa, O.H., Candemir, S., Poostchi, M., Yang, F., Kim, L., Ding, M., Folio, L.R., Antani, S., Gabrielian, A., et al. (2018). Detecting drug-resistant tuberculosis in chest radiographs. *Int. J. Comput. Assist. Radiol. Surg.* 13, 1915–1925.
  38. Ali, M.H., Khan, D.M., Jamal, K., Ahmad, Z., Manzoor, S., and Khan, Z. (2021). Prediction of multidrug-resistant tuberculosis using machine learning algorithms in SWAT, Pakistan. *J. Healthc. Eng.* 2021, 2567080.
  39. Li, Y., Wang, B., Wen, L., Li, H., He, F., Wu, J., Gao, S., and Hou, D. (2023). Machine learning and radiomics for the prediction of multidrug resistance in cavitary pulmonary tuberculosis: a multicentre study. *Eur. Radiol.* 33, 391–400.
  40. Zhou, W., Cheng, G., Zhang, Z., Zhu, L., Jaeger, S., Lure, F.Y.M., and Guo, L. (2022). Deep learning-based pulmonary tuberculosis automated detection on chest radiography: large-scale independent testing. *Quant. Imag. Med. Surg.* 12, 2344–2355.
  41. Linh, N.N., Viney, K., Gegia, M., Falzon, D., Glaziou, P., Floyd, K., Timimi, H., Ismail, N., Zignol, M., Kasaeva, T., and Mirzayev, F. (2021). World Health Organization treatment outcome definitions for tuberculosis: 2021 update. *Eur. Respir. J.* 58, 2100804.
  42. Yu, C., Gao, C., Wang, J., Yu, G., Shen, C., and Sang, N. (2021). BiSeNet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *Int. J. Comput. Vis.* 129, 3051–3068.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
BiSeNetV2	Yu et al. <sup>41</sup> (2021)	<a href="https://doi.org/10.1007/s11263-021-01515-2">https://doi.org/10.1007/s11263-021-01515-2</a>
Gated Recurrent Unit (GRU)	Chung et al. <sup>42</sup> (2014)	<a href="https://doi.org/10.48550/arXiv.1412.3555">https://doi.org/10.48550/arXiv.1412.3555</a>
Prediction models (PMs)	This study	Please request from lead contact ( <a href="mailto:zxgks@163.com">zxgks@163.com</a> ) for non-commercial, research purposes
Matplotlib	Version 3.3.1	<a href="https://matplotlib.org/3.3.1/">https://matplotlib.org/3.3.1/</a>
Scikit-learn	Version 0.23.2	<a href="https://scikit-learn.org/stable/whats_new/v0.23">https://scikit-learn.org/stable/whats_new/v0.23</a>
Python	Version 3.80	<a href="https://www.python.org/downloads/release/python-380/">https://www.python.org/downloads/release/python-380/</a>
SPSS	Version 20	<a href="https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-20">https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-20</a>

## RESOURCE AVAILABILITY

## Lead contact

Further information and requests should be directed to the lead contact, Xiaoguang Zou ([zxgks@163.com](mailto:zxgks@163.com)).

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

- All data reported in this paper will be shared by the [lead contact](#) upon request.
- All original code has been deposited at github and is publicly available as of the date of publication. DOI is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

## 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]KDYIRB(No.99)) and the patient consent was waived.

## METHOD DETAILS

## 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. Patients with bacteriologically confirmed TB were collected from two local hospitals in China, named Site A and Site B. With the criteria of data inclusion and exclusion, a total of 579 eligible patients derived from the original 2,013 patients were ultimately included, and divided into internal and external datasets (Figure 4). Site A was collected between January 2020 and September 2021 to develop the AI models. It contained 493 patients with 51.2% of male and a median age of  $58.66 \pm 18.66$  years. The training and testing of the models were conducted at an 8:2 ratio. Site B contained 86 patients who were collected between January 2021 and September 2021 for independent external testing. Among them, 40.7% were male and the median age was  $58.85 \pm 17.80$  years. The criteria for exclusion were determined as (i) TB patients confirmed by Acid-fast bacilli culture positivity with presence 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 at the second and the sixth months; (iv) Three types of treatment outcomes were included: cured, treatment completed and DR-TB.

In this study, a total of 579 eligible patients were enrolled and treated with a standard 6-month 2HRZ/4HR regimen. The treatment involved a 2-month intensive phase of daily isoniazid, rifampicin, and pyrazinamide, followed by a 4-month continuation phase of daily isoniazid and rifampicin. At the end of the second month of treatment, all patients underwent a CT scan and sputum culture examination. If the sputum

culture remained positive, drug susceptibility testing was performed to determine if the patient was drug-sensitive or drug-resistant. If the patient was found to be drug-sensitive, the treatment regimen remained the same, but an additional sputum examination was conducted at the end of the third month. However, 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 6-month treatment regimen, all patients underwent another CT scan. The treatment outcomes were classified as three TB treatment outcomes according to the guideline of WHO (Table 3).<sup>41</sup> Three types of treatment outcomes were included: cured, treatment completed and DR-TB. Cure and treatment completion were considered successful TB treatment, and the DR-TB who acquired drug resistance during the course of treatment was considered treatment failure in the study.

### CT image acquisition and image preprocessing

The first prediction model (PM1) was developed using only the pretreatment CT scans. Follow-up scans taken at the end of the second and sixth months were added to the first model in sequence to create the second and third prediction models (PM2 and PM3, respectively).

In the study, CT scans were acquired according to standardized protocols at each hospital. The internal cohort underwent CT scans using two machines: 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). For the external cohort, CT scans were performed using the 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). To ensure patient privacy, all identification information was removed from the imaging data before preprocessing. We first developed a pre-trained model using all CT scans collected from Site A, and then the patients who had undergone three or more CT scans were involved for the prediction model construction where they were divided into a training set and an internal testing set (Figure 4). The first prediction model (PM1) was developed using only the pretreatment CT scans. Follow-up scans taken at the end of the second and sixth months were added to the first model in sequence to create the second and third prediction models (PM2 and PM3, respectively).

During the labeling process, two radiologists who had more than 10 years of experience in interpreting CT images marked the TB lesions on each slice of CT scans to produce a Dice coefficient value independently, which is a measure of the similarity between two sets of data. If the two Dice values were both  $\geq 0.95$ , the averaged value would be used as the ground truth of the image. Otherwise, a senior radiologist with more than 30 years of experience would be involved to make the final determination. To improve the performance of the model, we used image augmentation techniques including flipping, translation, rotation, and deformation on the internal dataset. Image augmentation is a method of increasing the size and diversity of a dataset by creating modified versions of the original images while preserving their underlying characteristics.<sup>40</sup>

### Development of the PM1, PM2 and PM3

A pretrained model was developed first to predict the areas suspected of tuberculosis by extracting and learning features from manual lesions. Then the Gated Recurrent Unit (GRU) was used to classify the predicted feature map. To balance the speed and accuracy of the pre-trained model, a bilateral segmentation network (BiSeNetV2-3D)<sup>42</sup> was proposed, and the architecture of the network is shown in Figure 5. It includes three main components: (1) A detail branch where wide channels and shallow layers were involved was used to capture low-level details and generate high-resolution feature representations; (2) A lightweight semantic branch characterized narrow channels and deep layers was used to obtain high-level semantic context. (3) A guiding aggregation layer was designed to enhance the two types of feature representation: interconnection and fusion. BiSeNetV2-3D achieves this balance by dividing the image into two parts, a low-resolution sub-image and a high-resolution sub-image, and processing them separately in parallel. The low-resolution sub-image is processed by the lightweight semantic branch to capture the high-level semantic context, while the high-resolution sub-image is processed by the detail branch to capture the low-level details and generate high-resolution feature representations. The output features from the two branches are then aggregated using the guiding aggregation layer, which enhances both interconnection and fusion between the features. This approach allows the model to balance the trade-off between accuracy and speed, as the lightweight semantic branch is faster but less accurate, while the detail branch is more accurate but slower. By processing the image in parallel using both branches and then aggregating the features, the BiSeNetV2-3D can achieve a high level of accuracy while still maintaining a fast-processing speed. This makes it a suitable architecture for our prediction task, where only small and limited dataset were involved and both accuracy and speed are crucial for the model.

In detail, the detail branch was a shallow structure with a small span whose key point was to use wide channels and shallow layers to deal with spatial details. The instantiation of detail branch included three stages, each layer is convolution layer, followed by batch normalization and activation function. The first layer of each stage had a stride of 2, while other layers of the same stage had the same number of convolution layers and the same output feature map sizes. Therefore, the output feature map extracted by this branch was 1/8 of the original input. Semantic branch paralleled with detail branch was used to capture advanced semantics and it includes Stem Block, Context Embedding Block and Gather-and-Expansion Layer. Two different down-sampling methods were used in the Stem Block to reduce the feature representation, and the output characteristics of the two branches were connected in series as the output. The Context Embedding Block was designed to expand the acceptance domain by applying global average pool and residual connection. For the Gather-and-Expansion Layer, a  $3 \times 3 \times 3$  convolution was involved to aggregate the characteristic responses and extend them to high-dimensional space; then  $3 \times 3 \times 3$  depth convolution is independently performed on each individual output channel of the expansion layer; finally, we used the  $1 \times 1 \times 1$  convolution as the projection layer, and the output of the deep convolution was projected into the low channel capacity space. Given that the feature

representations of detail branch and semantic branch are complementary, we designed a bilateral-guided aggregation layer to fuse these two types of feature representations. After identifying the region of interest (ROI), the GRU 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.

### **QUANTIFICATION AND STATISTICAL ANALYSIS**

Statistical analyses were performed using Python 3.80 and SPSS 20. The performance of the proposed prediction models of PM1, PM2 and PM3 to predict treatment outcomes on the internal and external dataset were assessed by the receiver operating characteristic curve (ROC) plotted by matplotlib and Scikit-learn. Besides, the AUC, accuracy, sensitivity, specificity and F1 score were also calculated.