Original Article

Differentiating between drug-sensitive and drug-resistant tuberculosis with machine learning for clinical and radiological features

Feng Yang[1]^, Hang Yu[1], Karthik Kantipudi[2], Manohar Karki[1], Yasmin M. Kassim[1], Alex Rosenthal[2], Darrell E. Hurt[2], Ziv Yaniv[2], Stefan Jaeger[1]

[1]Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA; [2]Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

Contributions: (I) Conception and design: F Yang, Z Yaniv, S Jaeger; (II) Administrative support: A Rosenthal, DE Hurt, S Jaeger; (III) Provision of study materials or patients: K Kantipudi, Z Yaniv, A Rosenthal, DE Hurt; (IV) Collection and assembly of data: K Kantipudi, Z Yaniv, A Rosenthal, DE Hurt; (V) Data analysis and interpretation: F Yang, H Yu, K Kantipudi, M Karki, YM Kassim; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Feng Yang; Dr. Stefan Jaeger. Lister Hill National Center of Biomedical Communications, National Library of Medicine, NIH, Bethesda, MD 20894, USA. Email: feng.yang2@nih.gov; stefan.jaeger@nih.gov.

Background: Tuberculosis (TB) drug resistance is a worldwide public health problem that threatens progress made in TB care and control. Early detection of drug resistance is important for disease control, with discrimination between drug-resistant TB (DR-TB) and drug-sensitive TB (DS-TB) still being an open problem. The objective of this work is to investigate the relevance of readily available clinical data and data derived from chest X-rays (CXRs) in DR-TB prediction and to investigate the possibility of applying machine learning techniques to selected clinical and radiological features for discrimination between DR-TB and DS-TB. We hypothesize that the number of sextants affected by abnormalities such as nodule, cavity, collapse and infiltrate may serve as a radiological feature for DR-TB identification, and that both clinical and radiological features are important factors for machine classification of DR-TB and DS-TB.

Methods: We use data from the NIAID TB Portals program (https://tbportals.niaid.nih.gov), 1,455 DR-TB cases and 782 DS-TB cases from 11 countries. We first select three clinical features and 26 radiological features from the dataset. Then, we perform Pearson's chi-squared test to analyze the significance of the selected clinical and radiological features. Finally, we train machine classifiers based on different features and evaluate their ability to differentiate between DR-TB and DS-TB.

Results: Pearson's chi-squared test shows that two clinical features and 23 radiological features are statistically significant regarding DR-TB *vs.* DS-TB. A ten-fold cross-validation using a support vector machine shows that automatic discrimination between DR-TB and DS-TB achieves an average accuracy of 72.34% and an average AUC value of 78.42%, when combing all 25 statistically significant features.

Conclusions: Our study suggests that the number of affected lung sextants can be used for predicting DR-TB, and that automatic discrimination between DR-TB and DS-TB is possible, with a combination of clinical features and radiological features providing the best performance.

Keywords: Differential diagnosis; tuberculosis (TB); drug-resistance (DR); clinical features; radiological features; machine learning

^ ORCID: 0000-0002-8334-7450.

Submitted Mar 16, 2021. Accepted for publication Jul 23, 2021.
doi: 10.21037/qims-21-290
View this article at: https://dx.doi.org/10.21037/qims-21-290

Introduction

Tuberculosis (TB) drug resistance is a global public health concern since it threatens the progress made in TB care and control (1). In 2019, there were an estimated 10 million new TB cases; approximately half a million cases are resistant to rifampicin, of which 78% are multidrug-resistant TB (MDR-TB) (2). MDR-TB is a type of TB that is resistant to at least two first-line anti-TB drugs: isoniazid and rifampicin. Drug-resistant TB is a growing public health concern since it requires more complex treatment than drug-sensitive TB and incurs more costs. In 2019, it was estimated that globally 3.3% of new TB cases and 17.7% of previously treated TB cases are MDR-TB (2).

Early identification of drug resistance enables patient-specific drug treatment, which reduces the period of infectiousness and disease spread in addition to improving outcomes. However, discrimination between drug-resistant TB (DR-TB) and drug-sensitive cases (DS-TB) using readily available clinical information and images, preferably during the first visit, is still an open problem. Currently, there are two types of TB drug susceptibility tests: conventional culture-based phenotypic testing and molecular testing. The former involves looking at the bacteria behavior, which requires a well-equipped laboratory facility and may take several weeks to obtain the results (3). The latter involves looking at genetic mutations, which is fast but expensive and may produce inconclusive results (4). Therefore, it is desirable to predict the suspicion of DR-TB automatically from radiological findings and clinical information in patient medical records.

Inspired by the work in (5), we hypothesize that the number of affected sextants may serve as important radiological features for DR-TB identification, and that both clinical and radiological features are important factors for machine classification between DR-TB and DS-TB.

Previous work

There is evidence that certain clinical data and radiological findings may enable differentiation between DR-TB and DS-TB. Faustini *et al.* (6) conducted a review of twenty-nine studies before 2006 and reported that prior treatment for TB is the strongest determinant of MDR-TB in Europe. Tembo and Malangu (7) collected 2,568 medical records in Botswana and found that previous treatment and positive sputum smear microscopy are associated with the prevalence of MDR-TB or rifampicin-resistant TB (RR-TB). Mdivani *et al.* (8) reported three risk factors for MDR-TB based on analysis of 1,422 patients from Georgia: retreatment case, history of injection drug use and female gender. O'Donnell *et al.* (9) found that women admitted to hospital in KwaZulu-Natal, South Africa with drug-resistant TB are 38% more likely than men to have XDR-TB based on an analysis of 4,514 patients. Shen *et al.* (10) conducted an analysis of 8419 patients from Shanghai, China, and stated that patients aged 30–59 years are more likely to be DR-TB in previously treated cases. Lv *et al.* (11) reported from 3,552 patients in Dalian, China, that previously treated patients and older age are more likely to have MDR-TB. Icksan *et al.* (12) compared chest X-ray (CXR) findings from 183 MDR-TB and 183 DS-TB cases in Indonesia and reported that the MDR-TB group has more large-size lesions while the DS-TB group has more small- and medium-size lesions. Wáng *et al.* (13) performed a review on available articles before 2018 for radiological signs of MDR-TB and found that thick-walled multiple cavities (particularly with count ≥3 and size ≥30 mm) present the most promising radiological sign for MDR-TB with good specificity but at the cost of low sensitivity. Huang *et al.* (5) reported from 468 DR-TB cases and 223 DS-TB cases that a combination of consolidated nodule number and size can be used to predict the probability of MDR-TB. Flores-Treviño *et al.* (14) found from 144 patients in Mexico that multiple cavities is a predictor for DR-TB.

Based on these past studies, we hypothesized that the number of affected sextants may serve as a useful feature when applying machine learning techniques to the discrimination of DR-TB and DS-TB. To date, very few works have been concerned with discriminating between DR-TB and DS-TB in an automated manner. Kovalev *et al.* (15) apply different machine learning methods to features extracted from CXR images or CT images or both. They achieve an accuracy of 73% with an AUC value of 72%, a sensitivity of 82% and a specificity of 58% when combining the features from both X-ray and CT images. The accuracy

Table 1 Origin of the TB cases analyzed in this work

Country	No. (%) of cases		
	DS-TB	DR-TB	Total
Azerbaijan	0 (0)	7 (0.31)	7 (0.31)
Belarus	113 (5.05)	461 (20.61)	574 (25.66)
Georgia	364 (16.27)	340 (15.20)	704 (31.47)
India	151 (6.75)	49 (2.19)	200 (8.94)
Kazakhstan	35 (1.56)	223 (9.97)	258 (11.53)
Kyrgyzstan	0 (0)	110 (4.92)	110 (4.92)
Moldova	1 (0.04)	26 (1.16)	27 (1.21)
Republic of the Congo	0 (0)	1 (0.04)	1 (0.04)
Romania	4 (0.18)	87 (3.89)	91 (4.07)
South Africa	94 (4.20)	3 (0.13)	97 (4.34)
Ukraine	20 (0.89)	148 (6.62)	168 (7.51)

for CXR images alone or CT images alone falls to 62% and 65%, respectively. Our previous work (16) applied both traditional machine-learning methods and deep learning networks to CXR images, achieving an AUC value around 66% and an accuracy around 60%.

Methods

Data collection

We use a dataset of 2,237 patients, which includes de-identified clinical data and CXR images publicly available from the NIAID TB Portals program (17). Each patient record is manually annotated with clinical information and radiological findings based on CXR images. Clinical information includes age of onset, gender, patient type (new, relapse or failure), type of sample (pulmonary or extrapulmonary), BMI, diagnosis, prescription drugs, laboratory tests, treatment period, treatment status and outcome. A new case refers to a patient who has never been treated for TB or has taken anti-TB drugs for less than one month. A relapse case refers to a patient who has previously been treated for TB, was declared cured or completed treatment at the end of the most recent course of treatment, and is now diagnosed with a recurrent episode of TB (either a true relapse or a new episode of TB caused by reinfection). A failure case represents a patient who has previously been treated for TB and whose treatment failed at the end of the most recent course of treatment (17). Radiological findings include chest radiography patterns such as the number and location of affected sextants, the presence of mediastinal lymphadenopathy, presence of other non-TB abnormalities, the overall percentage of abnormal volume, and the pleural effusion percentage of the hemithorax involved. Due to financial constraints and the size of the TB portals CXR dataset, radiological features are obtained using a single experienced radiologist-reading per image. The whole dataset was annotated by multiple radiologists from the countries contributing data to the program. Due to the large number of radiologists participating in this study, their annotations are not biased toward a single radiologist. The 2,237 patients include 782 DS-TB and 1,455 DR-TB patients, acquired from 11 countries. Distribution of the origin is listed in *Table 1*. The type of drug susceptibility was determined by sputum cultured drug resistance testing and/or molecular testing, specifically Bactec, Hain (FL-LPA), SL-LPA, GeneXpert, and Lowenstein -Jensen testing (17,18). Data usage is exempt from local institutional review board review as it is publicly available from the TB portals program. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The TB portals program participants are responsible for ensuring compliance with their countries' laws, regulations, and ethics considerations (17).

Statistical analysis

In this work, we hypothesize that the number of affected sextants may serve as an important radiological feature for DR-TB identification, and that both clinical and radiological features are relevant for machine classification of DR-TB and DS-TB. Lung sextants are defined by dividing each lung into three equal sections from apex to base, as shown in *Figure 1*. A sextant is said to be affected if either nodules, cavities, collapses, or infiltrates are present. We utilize three clinical features and 26 radiological features, as listed in *Table 2*. Age in this paper means age of onset. To gain insight into the statistical significance of extracted features with different types of resistance (DR-TB or DS-TB), Pearson's chi-squared test is applied. A feature with P<0.05 is considered statistically significant.

Machine classification of drug-sensitive and drug-resistant TB

Based on the clinical and radiological features selected by Chi-squared tests, we train a machine learning classifier,

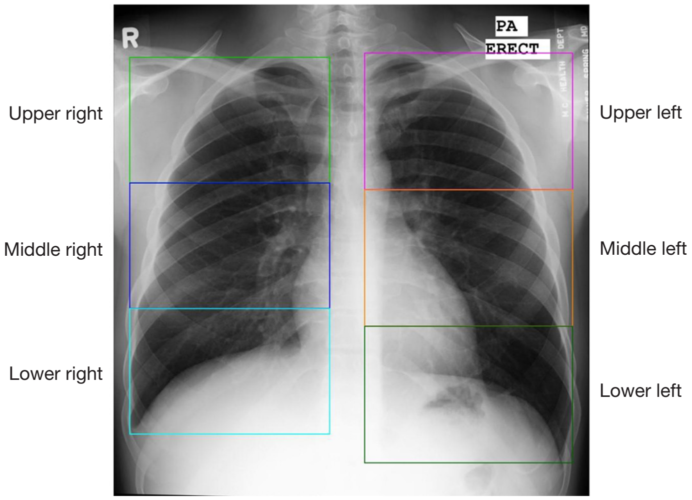


Figure 1 Definition of lung sextants in our study.

Table 2 Summary of clinical features and radiological features extracted from medical records

Type and number of features	Description
Three clinical features	Age; gender; patient type (new case, relapse, or failure)
Radiological features	
12 nodule features	Number of sextants affected by (I) either small nodules (<3 mm), (II) medium nodules (3–8 mm), (III) large nodules (8–30 mm), (IV) huge nodules (≥30 mm), (V) multiple nodules, (VI) calcified or partially calcified nodules, (VII) non-calcified nodules, (VIII) clustered nodules, (IX) low ground glass density active fresh nodules, (X) medium density stabilized fibrotic nodules, (XI) high density calcified typically sequellae nodules, or (XII) any kind of nodules
Six cavity features	Number of sextants affected by either (I) small cavities (<10 mm), (II) medium cavities (10–25 mm), (III) large cavities (>25 mm), (IV) multi-sextant cavities, (V) visible multiple cavities; or (VI) any kind of cavities
Eight other lung abnormality features	(I) Overall percentage of abnormal volume; (II) pleural effusion percentage of involved hemithorax; (III) number of sextants affected by collapse; (IV) number of sextants affected by low ground glass density infiltrates; (V) number of sextants affected by medium density infiltrates; (VI) number of sextants affected by high density infiltrates; (VII) presence of mediastinal lymphadenopathy; (VIII) presence of other non-TB abnormalities

Age in our work means age of onset.

a Support Vector Machine (SVM) (19), to discriminate between DS-TB and DR-TB. We illustrate the pipeline of our machine classification in *Figure 2*. To compare the contributions of different features for classifying DR-TB *vs.* DS-TB, we train the SVM classifier using different feature combinations.

Our dataset includes 782 DS-TB cases and 1,455 DR-TB cases, and is thus biased toward DR-TB. Machine learning classifiers are sensitive to the proportions of different classes in the training set. If ignored, data imbalance will bias predictions in favor of the majority class, leading to inaccurate results. To balance the dataset, several approaches can be applied: down-sampling of the majority class, over-sampling of the minority class, or synthetic minority over-sampling (SMOTE) type techniques (20-23). Data down-sampling involves randomly removing samples from the majority class, which might discard useful information. Data over-sampling is a process of randomly duplicating samples of the minority class, which will not lose any information from the original dataset, but is prone to over-fitting to the training data. The SMOTE technique is an improved over-sampling method that synthesizes new samples from

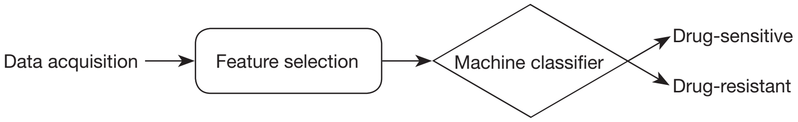
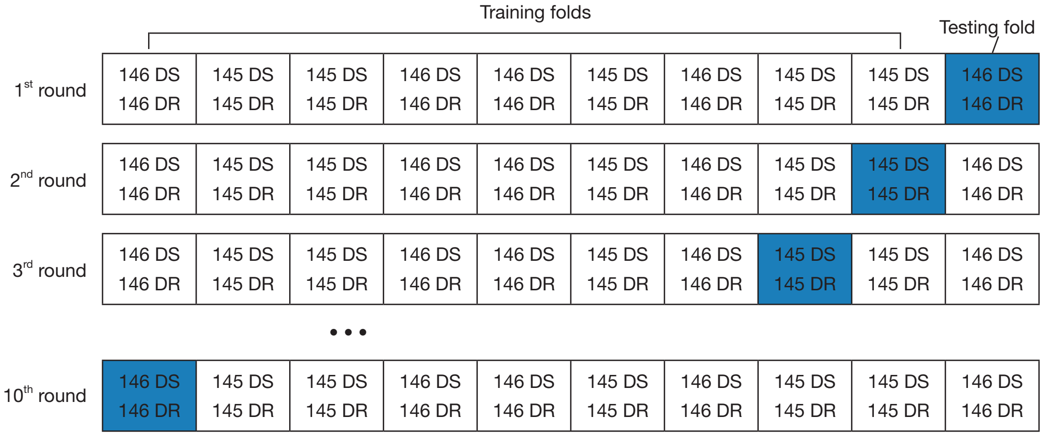


Figure 2 Workflow of the proposed machine classification between DR-TB and DS-TB. DS-TB, drug-sensitive tuberculosis; DR-TB, drug-resistant tuberculosis.



1st round	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR	146 DS 146 DR	145 DS 145 DR	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR
2nd round	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR	146 DS 146 DR	145 DS 145 DR	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR
3rd round	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR	146 DS 146 DR	145 DS 145 DR	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR
•••										
10th round	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR	146 DS 146 DR	145 DS 145 DR	146 DS 146 DR	145 DS 145 DR	145 DS 145 DR	146 DS 146 DR

Figure 3 Ten-fold cross validation based on balanced data with oversampled DS-TB cases. DS-TB, drug-sensitive tuberculosis.

the minority class. It often outperforms simple/random over-sampling (21-23). In our study, we use the classic SMOTE method (20) for data over-sampling in imbalanced binary dataset classification. We evaluate the classification performance based on ten-fold cross validation: 782 DS-TB cases and 1,455 DR-TB cases are first divided into ten folds, each of which includes 78 (or 79) DS-TB cases and 145 (or 146) DS-TB cases. To balance the data between DS-TB and DR-TB cases, the 78 (or 79) DS-TB cases in each fold are oversampled to 145 (or 146) cases using the SMOTE method, or the 145 (or 146) DR-TB cases in each fold are down-sampled to 78 (or 79). Then, nine folds are used for training data and the remaining fold is used as testing data to calculate the accuracy. The process is repeated ten times, with each fold serving once as testing data, and the average performance of the ten rounds is reported as the final evaluation result. *Figure 3* visually illustrates the ten-fold cross validation scheme on balanced data obtained via the SMOTE oversampling method.

Results

Figure 4 illustrates the distribution of the three clinical features we use (patient type, age and gender) for the 1,455 DR-TB and 782 DS-TB cases analyzed in this work. We find visible differences for patient type distributions between DR-TB and DS-TB: For DR-TB cases, 62.96% are in the *New* category, 23.71% are in the *Relapse* category, and 13.33% are in the *Failure* category. However, for DS-TB cases, 90.41% are in the *New* category, 9.34% are in the *Relapse* category, and only 0.26% are in the *Failure* category. Age categories follow those defined in (24), with a slightly finer resolution. Clear differences are observed among age groups between DR-TB and DS-TB. Both DS-TB and DR-TB patients are more likely to be less than 65 years old, whereas the frequency of DR-TB is higher in the age groups of 35–44 and 45–54. No clear difference is observed for the gender distribution between DR-TB and DS-TB. Differences observed in feature distributions are consistent with the statistical significance analysis using Pearson's chi-squared test, which is performed based on the null hypothesis that the clinical feature categories are independent from being drug-sensitive or drug-resistant. Patient type (P<0.001) and age (P<0.01) show a statistically significant association with resistance type. More details about the distribution and chi-squared test results of the three clinical features can be found in Table S1.

Figure 5 shows the distribution of 12 nodule features.

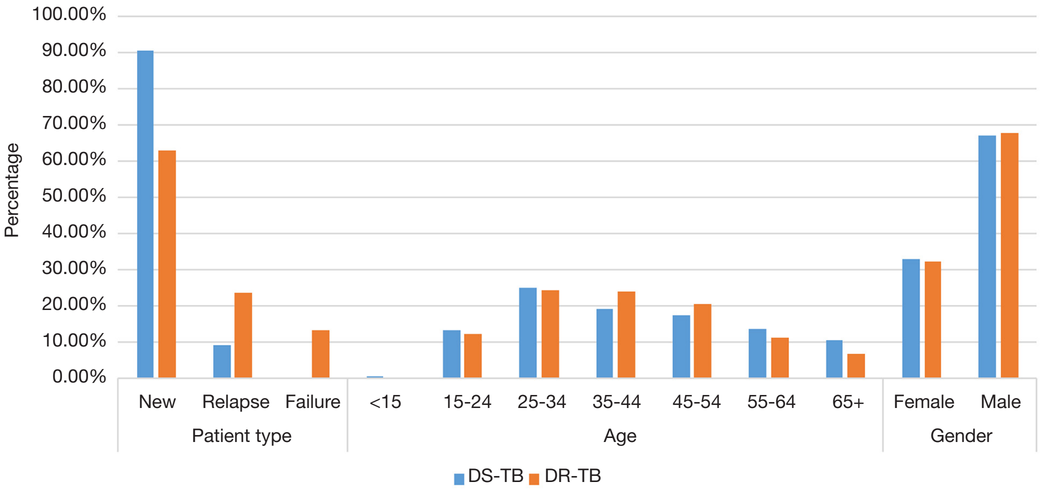


Figure 4 Distributions of three clinical features for 782 DS-TB patients (blue) and 1,455 DR-TB patients (orange). Shown are the percentage of cases for each feature present in a given category (e.g., the blue bars for patient type show that 90.41% of DS-TB patients are in the *New* category, 9.34% in the *Relapse* category, and 0.26% in the *Failure* category). DS-TB, drug-sensitive tuberculosis; DR-TB, drug-resistant tuberculosis.

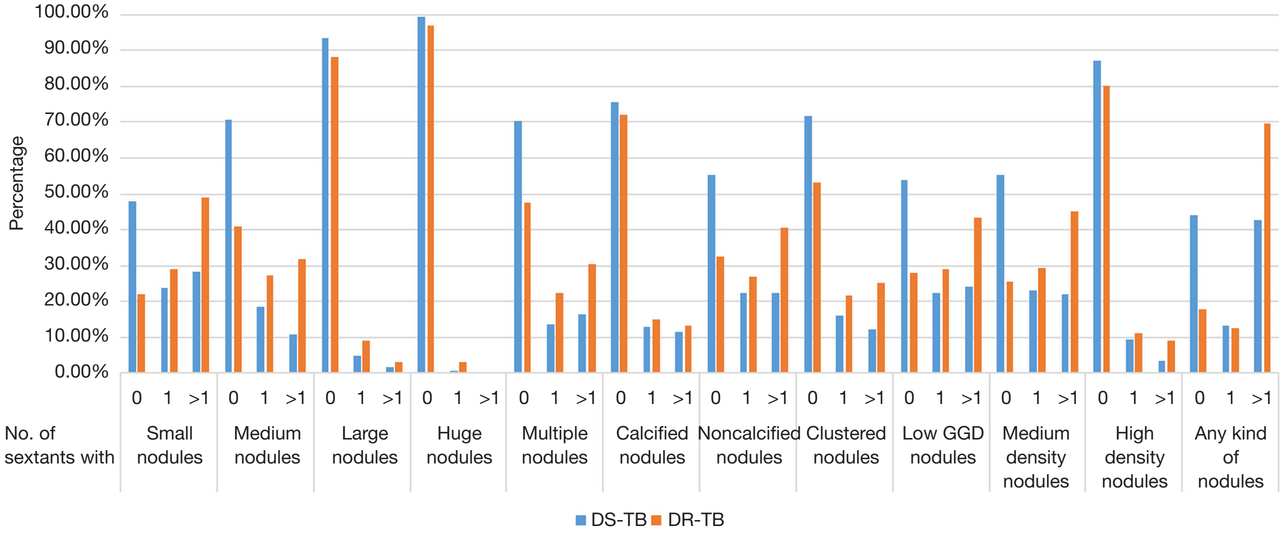


Figure 5 Distributions of 12 nodule related features for 782 DS-TB patients (blue) and 1,455 DR-TB patients (orange). Shown are the percentages of cases for each feature present in a given category (e.g., the blue bars in the category *Small nodules* show that 48.08% of DS-TB cases have no sextant affected by small nodules, 23.79% have only one sextant affected by small nodules, and 28.13% have more than one sextant affected by small nodules). DS-TB, drug-sensitive tuberculosis; DR-TB, drug-resistant tuberculosis.

Visible differences are observed between DS-TB and DR-TB. Nodules occur in approximately 82% of DR-TB patients and in approximately 56% of DS-TB patients. Multiple small or medium nodules, large nodules, huge nodules, multiple non-calcified nodules, multiple clustered nodules, multiple low ground-glass-density (GGD) nodules,

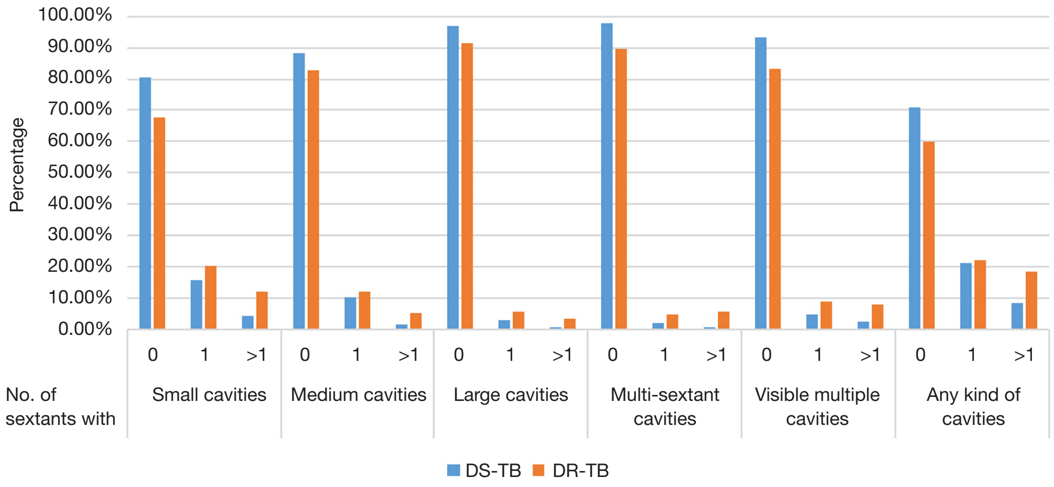


Figure 6 Distributions of six cavity-related features for 782 DS-TB patients (blue) and 1,455 DR-TB patients (orange). Shown are the percentages of cases for each feature present in a given category (e.g., the blue bars in the category *Small cavities* show that 80.31% of DS-TB cases have no sextant affected by small cavities, 15.60% have only one sextant affected by small cavities, and 4.09% have more than one sextant affected by small cavities). DS-TB, drug-sensitive tuberculosis; DR-TB, drug-resistant tuberculosis.

multiple medium-density nodules and multiple high-density nodules are more likely to occur in DR-TB. Pearson's chi-squared tests on the 12 nodule features show that all the nodule features are statistically significant regarding DR-TB *vs.* DS-TB, except for the number of sextants with calcified nodules (P>0.05). *Figure 6* depicts the distribution of six cavity features. We find that cavities occur in approximately 40% of DR-TB patients and in approximately 29% of DS-TB patients. Multiple small cavities, multiple medium cavities, and large cavities tend to occur in DR-TB patients. Pearson's chi-squared tests indicate that all six cavity features are statistically significant regarding DR-TB *vs.* DS-TB. *Figure 7* shows the distribution of eight other lung abnormality features. We observed that multiple sextants with collapses, multiple sextants with low GGD infiltrates, high-density infiltrates, more than 50% abnormal volume, and mediastinal lymphadenopathy are more likely to occur in DR-TB. Chi-squared test results show that, the number of sextants affected by medium density infiltrates (P>0.05) and the presence of other non-TB abnormalities (P>0.05), are not statistically significant with respect to discriminating between DR-TB and DS-TB. Differences visible in the feature distributions in *Figures 4-7* are consistent with Pearson's chi-squared tests. More details about distributions and chi-squared test results of the 26 radiological features can be found in Tables S2,S3.

To investigate the possibility of automatically differentiating between DR-TB and DS-TB and to evaluate the importance of specific features, we train SVM machine classifiers using eight different combinations of features listed in *Table 2*: (I) three clinical features; (II) 12 nodule features; (III) six cavity features; (IV) 12 nodule features plus six cavity features; (V) three clinical features plus 12 nodule features plus six cavity features; (VI) all 26 radiological features; and (VII) 25 significant clinical and radiological features. *Tables 3,4* show the results for machine classification of DR-TB cases *vs.* DS-TB cases using data down-sampling (*Table 3*) and data augmentation (*Table 4*) for data balancing, respectively. Comparing the results in *Tables 3,4*, we find that (I) classifiers with data augmentation achieve better results; (II) classifiers using only clinical features obtain an accuracy around 61%, with very low sensitivity; (III) compared to classifiers using only clinical features, classifiers using radiological features achieve a higher sensitivity at the cost of much lower specificity; (IV) classifiers using a combination of clinical and radiological features achieve a better performance than those using any of them alone; (V) the classifier using the 25 statistically significant features with SMOTE data augmentation achieves the best performance, with an average accuracy

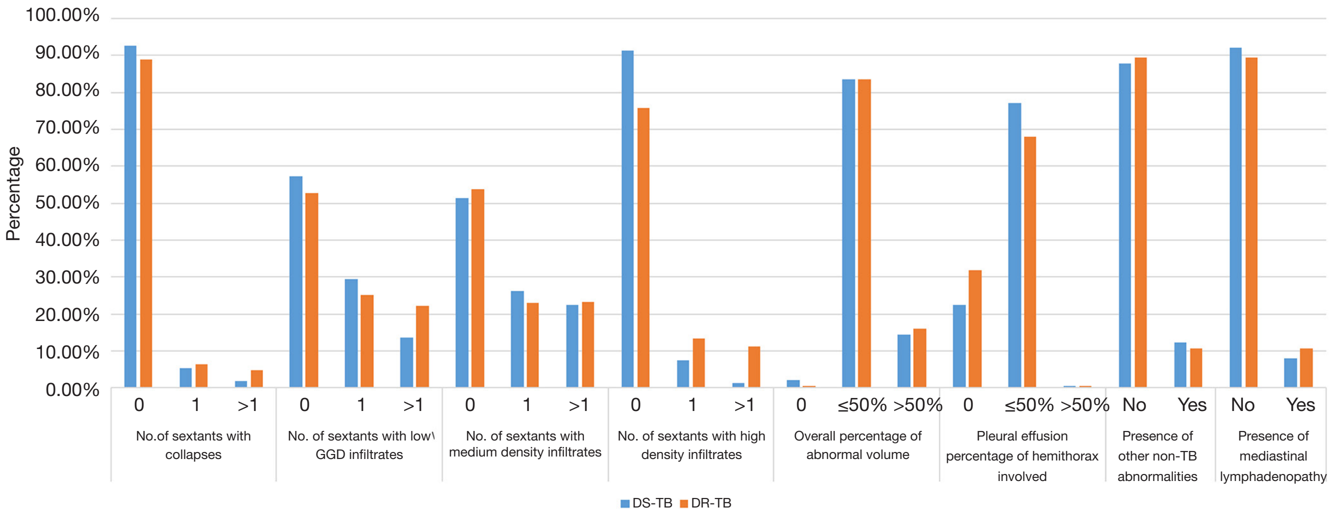


Figure 7 Distributions of eight other lung abnormality features for 782 DS-TB patients (blue) and 1,455 DR-TB patients (orange). Shown are the percentages of cases for each feature present in a given category (e.g., the blue bars in the category *No. of sextants with collapses* show that 92.71% of DS-TB cases have no sextant affected by collapse, 5.37% have only one sextant affected by collapse, and 1.92% have more than one sextant affected by collapse). DS-TB, drug-sensitive tuberculosis; DR-TB, drug-resistant tuberculosis.

Table 3 Ten-fold cross validation on balanced dataset with data down-sampling between DS-TB (782 cases) and DR-TB (782 cases)

Training features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC (%)
3 clinical features	60.22±3.05	40.04±7.54	84.40±4.84	72.33±5.57	60.77±4.95
12 nodule features	64.38±4.26	71.99±6.81	56.77±4.83	62.46±3.65	67.18±4.58
6 cavity features	55.62±3.08	30.55±5.70	80.69±5.60	61.69±7.51	55.03±3.29
12 nodule + 6 cavity features	64.32±4.82	71.99±6.66	56.64±5.98	62.43±4.25	67.38±4.96
3 clinical + 12 nodule + 6 cavity features	66.63±4.53	66.12±6.02	67.14±5.35	66.85±4.45	70.94±4.68
26 radiological features	66.44±4.87	*76.47±6.12*	56.40±5.50	63.70±3.97	71.13±4.68
3 clinical + 26 radiological features	68.22±4.77	73.01±5.06	*63.43±6.85*	*66.78±4.61*	*74.05±5.84*
25 significant features	*68.29±3.40*	75.70±5.49	60.87±5.91	66.04±3.38	73.60±4.79

Twenty-six radiological features include 12 nodule features, six cavity features and eight other lung abnormality features. Twenty-five significant features are obtained by excluding the four non-significant features from the 29 features. The best performance in each column is marked in italic.

of 72.34% and an average AUC value of 78.42%. The 25 significant features are obtained by excluding the four non-significant features (i.e., gender, number of sextans with calcified nodules, number of sextants with medium density infiltrates, and presence of other non-TB abnormalities) from the 29 features (three clinical features and 26 radiological features). The ROC curve for SVM-based machine classification using the 25 significant features is shown in *Figure 8*. More details of our trained SVM model can be found in Table S4.

Discussion

Limitation of the study

The current study has some limitations. First, we did not

Table 4 Ten-fold cross validation on balanced dataset with SMOTE data augmentation between DS-TB (1,455 cases) and DR-TB (1,455 cases)

Training features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC (%)
3 clinical features	61.89±3.31	38.43±5.27	85.35±7.11	73.63±9.26	64.44±3.07
12 nodule features	64.94±3.56	72.71±3.20	57.17±4.96	63.00±3.39	68.79±3.64
6 cavity features	56.77±2.43	34.02±3.34	79.53±5.64	63.02±6.29	56.60±3.76
12 nodule + 6 cavity features	65.33±4.03	70.38±3.64	60.28±6.42	64.08±4.16	68.76±4.42
3 clinical + 12 nodule + 6 cavity features	67.28±3.30	70.44±3.12	64.13±6.77	66.47±3.83	73.14±2.98
26 radiological features	67.28±5.35	84.13±2.81	56.95±13.34	67.40±7.46	72.25±6.24
3 clinical + 26 radiological features	70.99±3.18	74.36±2.84	67.63±7.36	69.98±4.42	77.56±3.72
25 significant features	*72.34±2.65*	*75.33±3.36*	*69.35±6.29*	*71.30±3.63*	*78.42±2.63*

Twenty-six radiological features include 12 nodule features, six cavity features and eight other lung abnormality features. Twenty-five significant features are obtained by excluding the four non-significant features from the 29 features. The best performance in each column is marked in italic.

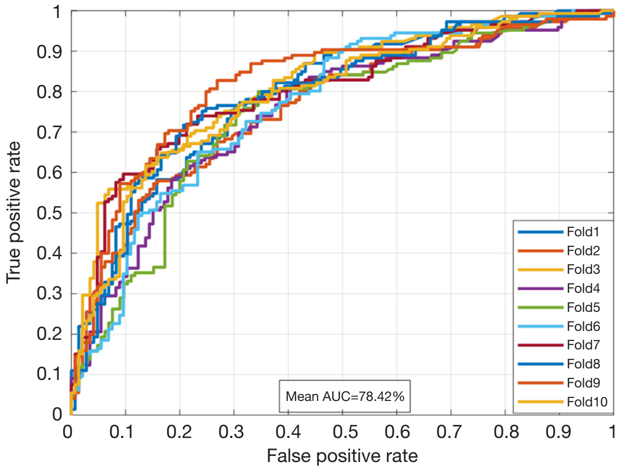


Figure 8 ROC curve of SVM classifier based on 25 significant features with data over-sampling and 10-fold cross validation. ROC, receiver operating characteristic; SVM, Support Vector Machine; AUC, area under the curve.

use radiological features from chest CT images, since many of the existing records do not include annotated CT radiological findings. Second, distribution bias is present in the source of both DR-TB and DS-TB. This work includes 2,237 patients from 11 countries. However, about 97% of all drug-sensitive cases are from five countries (Georgia, Belarus, India, South Africa, and Kazakhstan) and about 81% of all drug-resistant cases are from four countries (Belarus, Georgia, Kazakhstan and Ukraine). That is, our machine classifier learns drug-sensitive and drug-resistant features primarily from six countries, and the classification performance will decrease when we use it to identify DR-TB from other countries or when we perform a country-level evaluation. Third, DR-TB and DS-TB cases are imbalanced. We used SMOTE data augmentation techniques to balance the dataset; it would be better to have a truly balanced dataset for both the statistical significance analysis and machine classifier training. We are now working to increase the sample size of DR-TB and DS-TB cases to obtain a balanced data set with more uniformly distributed countries of origin. Fourth, our best machine learning model based on 25 significant features requires radiological findings reported by a radiologist, which limits the full automation of our machine learning model. Our next step will be to automatically detect such radiological features from radiographs based on deep learning methods.

Roles of patient type, age, and gender

In our study, the strongest predictor of DR-TB is patient type (New, Relapse or Failure). This is consistent with previous studies in Europe, Botswana, and Georgia (6-8). The treatment history is a well-known risk factor for the development of DR-TB. The WHO Global Report on Surveillance on MDR-TB and XDR-TB (25) stated that TB cases with a history of previous TB treatment are significantly associated with DR-TB. Such a predictor can be used in early screening of DR-TB cases, especially in resource-limited clinical settings. For example, a relapsed patient would indicate that drug susceptibility testing is recommended at treatment start.

The association between DR-TB and age is not well established in the previous reports since different studies use different cut-off points for age groups. The European MDR-TB analysis (6) indicated that MDR-TB patients are more likely to be younger than 65 years, while the report on MDR-TB in Shanghai, China (10) stated that the age group of 30–59 years is associated with MDR-TB. In our data, age also showed a significant association with DR-TB. However, we found both that DR-TB and DS-TB are more likely to happen in age groups less than 65 years, and that the frequency of DR-TB is higher than DS-TB in the age groups of 35–44 and 45–54.

Tuberculosis is more common in males (26-28). In (6) it was reported that male gender is a risk factor for MDR-TB cases in Western Europe, but not in Eastern Europe. Contrary to that observation, it was found in Georgia, that female gender is significantly associated with MDR-TB (8,9). The authors in (8) assumed that this association is related to the fact that the majority of health care workers are females in Georgia. In our study, we did not find a significant association between gender and DR-TB. We hypothesize that gender is likely a regional risk factor for DR-TB, but does not present a general association in the TB data from 11 countries used in this work.

Roles of radiological features

Previous works (29-31) have revealed that active TB is likely to affect the upper lung regions exhibiting cavities, consolidations, and nodules, and to affect unilateral lung regions exhibiting pleural effusions. However, based on a review paper from 2018 (13) and our literature search in PubMed on Feb 4, 2021, only a small number of reports have mentioned the importance of lesion types and their locations in the development of DR-TB. Both the systematic review of radiological signs for MDR-TB before 2018 (13) and the report on MDR-TB in Mexico (14) found that multiple cavities is a promising sign for identifying MDR-TB. This radiological feature may offer good specificity at the cost of low sensitivity (13). Our work confirms that there is a significant association between DR-TB and multiple cavities (P<0.001), and also quantitatively demonstrates that a SVM classifier using cavity lesions can predict DR-TB with an average specificity of 80%, at the cost of an average sensitivity of 34%. It was reported that MDR-TB patients are more likely to have large-size lesions, and DS-TB patients are more likely to have small- or medium-size lesions (12). Our study confirms that large nodules and large cavities are more common in DR-TB. In addition, we found that multiple nodules and multiple cavities are more common in DR-TB, which confirms the analysis in (5,13).

In our study, we also found that DR-TB patients are more likely to have more abnormalities in all the six lung sextants. *Figure 9* shows the abnormality distribution in the six lung sextants for DR-TB and DS-TB patients, using an abnormality occurrence index (AOI) that is calculated by dividing the sum of abnormalities of all patients in a given group for a given sextant by the number of patients in this group and sextant. A higher index indicates a higher possibility of abnormality occurrence in that sextant. In our future work, we will investigate the possibility of incorporating abnormality location into features for automated identification of DR-TB.

Conclusions

In this paper, we investigated the possibility of using the number of affected sextants for drug resistance prediction and the possibility of applying machine learning to discriminate between drug-resistant TB and drug-sensitive TB by incorporating both clinical and radiological features. We found that, clinical features can predict DR-TB cases with an accuracy of around 61%, with a relatively low sensitivity, while radiological features based on the number of affected sextants can predict DR-TB cases with an accuracy of around 67%, with low specificity. The combination of clinical and radiological features improves these results. For the combined features, our machine classifier achieves an average accuracy of 72.34% and an average AUC value of 78.42%. Our study suggests that the number of affected sextants can be used for identifying drug-resistant TB, and that automatic discrimination between drug-resistant TB and drug-sensitive TB is possible by utilizing both clinical features and radiological features.

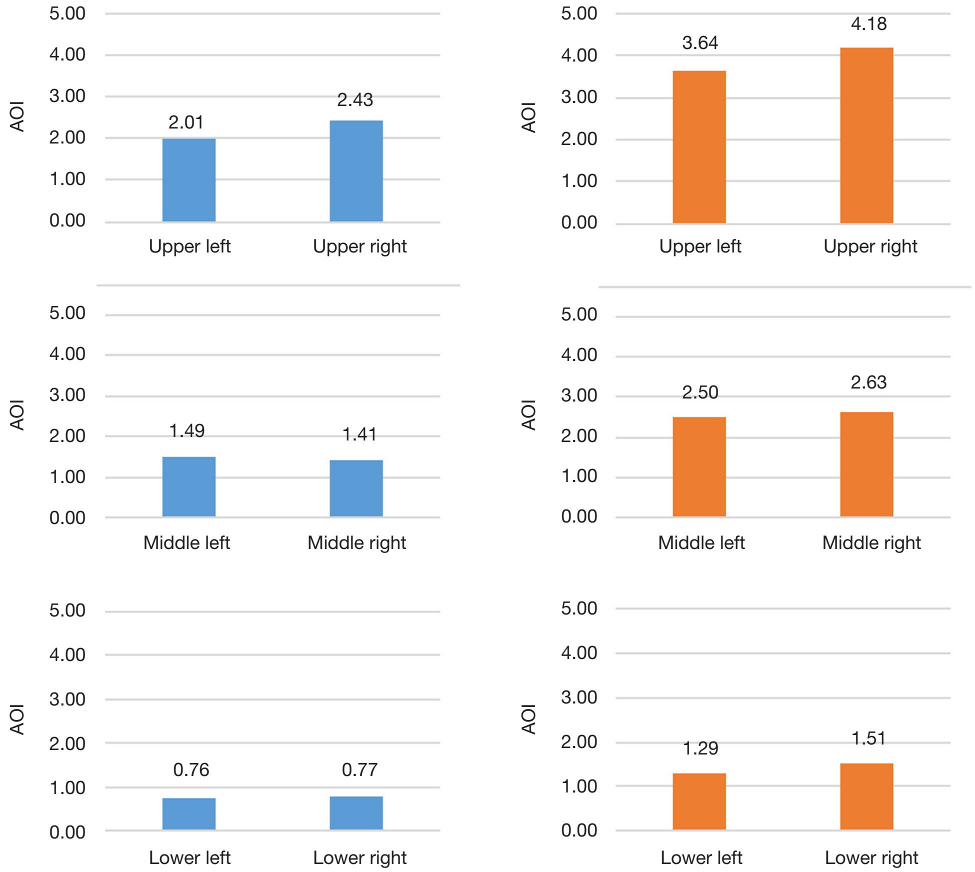


Figure 9 Abnormality distribution for 782 DS-TB patients (blue) and 1,455 DR-TB patients (orange). Shown are the abnormality occurrence indices (AOI) in each lung sextant (upper left, upper right, middle left, middle right, lower left and lower right). Taking DS-TB patients for example, the abnormality occurrence index in a given sextant is calculated by dividing the number of abnormalities in this sextant for all DS-TB patients by the number of DS-TB patients. A higher index indicates a higher possibility of abnormality occurrence in this sextant. DS-TB, drug-sensitive tuberculosis; DR-TB, drug-resistant tuberculosis.

Acknowledgments

Funding: This work was supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) under Interagency Agreement #750119PE080057, and by the Intramural Research Program of the National Library of Medicine, National Institutes of Health. It had been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases under BCBB Support Services Contract HHSN316201300006W/HHSN27200002. This research was also supported in part by an appointment to the National Library of Medicine Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the National Library of Medicine. ORISE is managed by ORAU under DOE contract number DE-SC0014664. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of NIH, NLM, DOE, or ORAU/ORISE.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/qims-21-290). Dr. SJ serves as an unpaid editorial board member of *Quantitative Imaging in Medicine and Surgery*. The other authors have no conflicts of interest to declare.

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). Data usage is exempt from local institutional review board review as it is publicly available from the TB portals program. The TB portals program participants are responsible for ensuring compliance with their countries' laws, regulations, and ethics considerations.

References

1. WHO. Module 1: Tuberculosis preventive treatment. Consolidated guidelines on tuberculosis 2020.
2. WHO. Global Tuberculosis Report 2020: Executive summary 2020.
3. World Health Organization (WHO). Technical report on critical concentrations for TB drug susceptibility testing of medicines used in the treatment of drug-resistant TB. WHO 2018.
4. Heyckendorf J, Andres S, Köser CU, Olaru ID, Schön T, Sturegård E, Beckert P, Schleusener V, Kohl TA, Hillemann D, Moradigaravand D, Parkhill J, Peacock SJ, Niemann S, Lange C, Merker M. What Is Resistance? Impact of Phenotypic versus Molecular Drug Resistance Testing on Therapy for Multi- and Extensively Drug-Resistant Tuberculosis. Antimicrob Agents Chemother 2018;62:e01550-17.
5. Huang XL, Skrahin A, Lu PX, Alexandru S, Crudu V, Astrovko A, et al. Prediction of multiple drug resistant pulmonary tuberculosis against drug sensitive pulmonary tuberculosis by CT nodular consolidation sign. bioRxiv [Internet] 2019; Available online: https://www.biorxiv.org/content/early/2019/11/07/833954
6. Faustini A, Hall AJ, Perucci CA. Risk factors for multidrug resistant tuberculosis in Europe: a systematic review. Thorax 2006;61:158-63.
7. Tembo BP, Malangu NG. Prevalence and factors associated with multidrug/rifampicin resistant tuberculosis among suspected drug resistant tuberculosis patients in Botswana. BMC Infect Dis 2019;19:779.
8. Mdivani N, Zangaladze E, Volkova N, Kourbatova E, Jibuti T, Shubladze N, Kutateladze T, Khechinashvili G, del Rio C, Salakaia A, Blumberg HM. High prevalence of multidrug-resistant tuberculosis in Georgia. Int J Infect Dis 2008;12:635-44.
9. O'Donnell MR, Zelnick J, Werner L, Master I, Loveday M, Horsburgh CR, Padayatchi N. Extensively drug-resistant tuberculosis in women, KwaZulu-Natal, South Africa. Emerg Infect Dis 2011;17:1942-5.
10. Shen X, DeRiemer K, Yuan ZA, Shen M, Xia Z, Gui X, Wang L, Gao Q, Mei J. Drug-resistant tuberculosis in Shanghai, China, 2000-2006: prevalence, trends and risk factors. Int J Tuberc Lung Dis 2009;13:253-9.
11. Lv XT, Lu XW, Shi XY, Zhou L. Prevalence and risk factors of multi-drug resistant tuberculosis in Dalian, China. In: Journal of International Medical Research 2017:1779-86.
12. Icksan AG, Napitupulu MRS, Nawas MA, Nurwidya F. Chest X-Ray Findings Comparison between Multi-drug-resistant Tuberculosis and Drug-sensitive Tuberculosis. J Nat Sci Biol Med 2018;9:42-6.
13. Wáng YXJ, Chung MJ, Skrahin A, Rosenthal A, Gabrielian A, Tartakovsky M. Radiological signs associated with pulmonary multi-drug resistant tuberculosis: an analysis of published evidences. Quant Imaging Med Surg 2018;8:161-73.
14. Flores-Treviño S, Rodríguez-Noriega E, Garza-González E, González-Díaz E, Esparza-Ahumada S, Escobedo-Sánchez R, Pérez-Gómez HR, León-Garnica G, Morfín-Otero R. Clinical predictors of drug-resistant tuberculosis in Mexico. PLoS One 2019;14:e0220946.
15. Kovalev V, Liauchuk V, Kalinovsky A, Rosenthal A, Gabrielian A, Skrahina A, Astrauko A, Tarasau A, Kalinouski A, Rosenthal A, Gabrielian A, Skrahina A, Astrauko A, Tarasau A. Utilizing radiological images for predicting drug resistance of lung tuberculosis. In: Computer Assisted Radiology and Surgery 2015:S129-30.
16. Jaeger S, Juarez-Espinosa OH, Candemir S, Poostchi M, Yang F, Kim L, Ding M, Folio LR, Antani S, Gabrielian A, Hurt D, Rosenthal A, Thoma G. Detecting drug-resistant tuberculosis in chest radiographs. Int J Comput Assist Radiol Surg 2018;13:1915-25.

17. Rosenthal A, Gabrielian A, Engle E, Hurt DE, Alexandru S, Crudu V, et al. The TB Portals: an Open-Access, Web-Based Platform for Global Drug-Resistant-Tuberculosis Data Sharing and Analysis. J Clin Microbiol 2017;55:3267-82.
18. Afsar I, Gunes M, Er H, Gamze Sener A. Comparison of culture, microscopic smear and molecular methods in diagnosis of tuberculosis. Rev Esp Quimioter 2018;31:435-8.
19. Cortes C, Vapnik V. Support-Vector Networks. Mach Learn 1995;20:273-97.
20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321-57.
21. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 2013;14:106.
22. Han H, Wang WY, Mao BH. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In: Lecture Notes in Computer Science 2005:878-87.
23. Nekooeimehr I, Lai-Yuen SK. Adaptive semi-unsupervised weighted oversampling (A-SUWO) for imbalanced datasets. Expert Syst Appl 2016;46:405-16.
24. Zhang X, Andersen AB, Lillebaek T, Kamper-Jørgensen Z, Thomsen VØ, Ladefoged K, Marrs CF, Zhang L, Yang Z. Effect of sex, age, and race on the clinical presentation of tuberculosis: a 15-year population-based study. Am J Trop Med Hyg 2011;85:285-90.
25. WHO. Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response [Internet]. Geneva: World Health Organization 2010. Available online: http://whqlibdoc.who.int/publications/2010/9789241599191_eng.pdf
26. Hertz D, Dibbern J, Eggers L, von Borstel L, Schneider BE. Increased male susceptibility to Mycobacterium tuberculosis infection is associated with smaller B cell follicles in the lungs. Sci Rep 2020;10:5142.
27. Neyrolles O, Quintana-Murci L. Sexual inequality in tuberculosis. PLoS Med 2009;6:e1000199.
28. Murphy ME, Wills GH, Murthy S, Louw C, Bateson ALC, Hunt RD, McHugh TD, Nunn AJ, Meredith SK, Mendel CM, Spigelman M, Crook AM, Gillespie SH; REMoxTB consortium. Gender differences in tuberculosis treatment outcomes: a post hoc analysis of the REMoxTB study. BMC Med 2018;16:189.
29. El-Solh AA, Hsiao CB, Goodnough S, Serghani J, Grant BJ. Predicting active pulmonary tuberculosis using an artificial neural network. Chest 1999;116:968-73.
30. Yeh JJ, Chen SC, Teng WB, Chou CH, Hsieh SP, Lee TL, Wu MT. Identifying the most infectious lesions in pulmonary tuberculosis by high-resolution multi-detector computed tomography. Eur Radiol 2010;20:2135-45.
31. Bhalla AS, Goyal A, Guleria R, Gupta AK. Chest tuberculosis: Radiological review and imaging recommendations. Indian J Radiol Imaging 2015;25:213-25.

Cite this article as: Yang F, Yu H, Kantipudi K, Karki M, Kassim YM, Rosenthal A, Hurt DE, Yaniv Z, Jaeger S. Differentiating between drug-sensitive and drug-resistant tuberculosis with machine learning for clinical and radiological features. Quant Imaging Med Surg 2022;12(1):675-687. doi: 10.21037/qims-21-290

Supplementary

Table S1 Distribution and chi-squared test results of clinical features for 782 DS-TB patients and 1,455 DR-TB patients. The null hypothesis of Pearson's chi-squared test is that clinical feature categories are independent from drug-sensitive or drug-resistant TB

Clinical features		No. (%) of cases		P value	Clinical features		No. (%) of cases		P value
		DS-TB	DR-TB				DS-TB	DR-TB	
Age	<15	5 (0.22)	3 (0.13)	1.40×10^{-3}	Patient type	New	707 (31.60)	916 (40.95)	5.64×10^{-46}
	15-24	105 (4.69)	182 (8.14)			Relapse	73 (3.26)	345 (15.42)	
	25-34	196 (8.76)	353 (15.78)			Failure	2 (0.09)	194 (8.67)	
	35-44	151 (6.75)	350 (15.65)						
	45-54	136 (6.08)	301 (13.46)		Gender	Female	257 (11.49)	468 (20.92)	0.74
	55-64	106 (4.74)	166 (7.42)			Male	525 (23.47)	987 (44.12)	
	65+	83 (3.71)	100 (4.47)						

Table S2 Distribution and chi-squared test results of 22 radiological features for 782 DS-TB patients and 1455 DR-TB patients with the null hypothesis that the number of affected sextants is independent from being DS-TB or DR-TB. GGD indicates ground glass density. The abbreviation "Multi" means multiple affected sextants

Radiological features	No. of affected sextants	No. (%) of cases		P value	Radiological features	No. of affected sextants	No. (%) of cases		P value
		DS-TB	DR-TB				DS-TB	DR-TB	
Small nodules	None	376 (16.81)	321 (14.35)	2.43×10^{-37}	Small cavities	None	628 (28.07)	988 (44.17)	8.72×10^{-12}
	Single	186 (8.31)	421 (18.82)			Single	122 (5.45)	292 (13.05)	
	Multi	220 (9.83)	713 (31.87)			Multi	32 (1.43)	175 (7.82)	
Middle nodules	None	553 (24.72)	594 (26.55)	1.88×10^{-43}	Medium cavities	None	692 (30.93)	1,207 (53.96)	9.24×10^{-6}
	Single	144 (6.44)	397 (17.75)			Single	80 (3.58)	173 (7.73)	
	Multi	85 (3.80)	464 (20.74)			Multi	10 (0.45)	75 (3.35)	
Large nodules	None	730 (32.6)	1,282 (57.31)	4.36×10^{-4}	Large cavities	None	759 (33.93)	1,329 (59.41)	2.14×10^{-7}
	Single	38 (1.70)	129 (5.77)			Single	21 (0.94)	80 (3.58)	
	Multi	14 (0.63)	44 (1.97)			Multi	2 (0.09)	46 (2.06)	
Huge nodules	None	776 (34.73)	1,405 (62.99)	5.20×10^{-4}	Multi-sextant cavities	None	764 (34.15)	1,304 (58.29)	3.93×10^{-12}
	Single	6 (0.27)	45 (2.01)			Single	16 (0.72)	69 (3.08)	
	Multi	0 (0)	5 (0.22)			Multi	2 (0.09)	82 (3.67)	
Multiple nodules	None	550 (24.59)	690 (30.84)	2.75×10^{-24}	Visible multiple cavities	None	729 (32.59)	1,214 (54.27)	1.96×10^{-10}
	Single	105 (4.69)	325 (14.53)			Single	35 (1.56)	127 (5.68)	
	Multi	127 (5.68)	440 (19.67)			Multi	18 (0.80)	114 (5.10)	
Calcified nodules	None	592 (26.46)	1,048 (46.85)	0.17	Any kind of cavities	None	553 (24.72)	870 (38.89)	1.34×10^{-10}
	Single	100 (4.47)	217 (9.70)			Single	166 (7.42)	320 (14.30)	
	Multi	90 (4.02)	190 (8.49)			Multi	63 (2.82)	265 (11.85)	
Noncalcified nodules	None	432 (19.31)	475 (21.23)	4.08×10^{-26}	Collapses	None	725 (32.41)	1,294 (57.85)	2.40×10^{-3}
	Single	175 (7.82)	391(17.48)			Single	42 (1.88)	93 (4.16)	
	Multi	175 (7.82)	589 (26.33)			Multi	15 (0.67)	68 (3.04)	
Clustered nodules	None	562 (25.12)	774 (34.60)	3.71×10^{-18}	Low GGD infiltrates	None	447 (19.98)	765 (34.20)	3.84×10^{-6}
	Single	126 (5.63)	315 (14.08)			Single	229 (10.24)	367 (16.41)	
	Multi	94 (4.20)	366 (16.36)			Multi	106 (4.74)	323 (14.44)	
Low GGD nodules	None	420 (18.78)	405 (18.10)	8.50×10^{-34}	Medium density infiltrates	None	401 (17.93)	781 (34.91)	0.23
	Single	174 (7.78)	420 (18.78)			Single	205 (9.16)	334 (14.93)	
	Multi	188 (8.40)	630 (28.16)			Multi	176 (7.87)	340 (15.20)	
Medium density nodules	None	432 (19.31)	373 (16.67)	2.55×10^{-45}	High density infiltrates	None	714 (31.92)	1,100 (49.17)	1.89×10^{-21}
	Single	179 (8.00)	428 (19.13)			Single	58 (2.59)	192 (8.58)	
	Multi	171 (7.64)	654 (29.24)			Multi	10 (0.45)	163 (7.29)	
High density nodules	None	682 (30.49)	1,164 (52.03)	1.12×10^{-6}	Any kind of nodules	None	345 (15.42)	258 (11.53)	1.47×10^{-42}
	Single	74 (3.31)	161 (7.20)			Single	103 (4.60)	184 (8.23)	
	Multi	26 (1.16)	130 (5.81)			Multi	334 (14.93)	1,013 (45.28)	

Table S3 Distribution and chi-squared test results of four radiological features for 782 DS-TB patients and 1,455 DR-TB patients based on the null hypothesis that the lesion volume or the presence of lesion is independent from DS-TB or DR-TB

Radiological features		No. (%) of cases		P value	Radiological features		No. (%) of cases		P value
		DS-TB	DR-TB				DS-TB	DR-TB	
Overall percentage of abnormal volume	0	17 (0.76)	8 (0.36)	1.50×10^{-3}	Pleural effusion percentage of hemithorax involved	0	175 (7.82)	461 (20.61)	1.98×10^{-5}
	≤50%	653 (29.19)	1,213 (54.22)			≤50%	603 (26.96)	987 (44.12)	
	>50%	112 (5.01)	234 (10.46)			>50%	4 (0.18)	7 (0.31)	
Presence of other non-TB abnormalities	No	686 (30.67)	1,299 (58.07)	0.27	Presence of mediastinal lymphadenopathy	No	720 (32.19)	1,301 (58.16)	0.04
	Yes	96 (4.29)	156 (6.97)			Yes	62 (2.77)	154 (6.88)	

Table S4 Detailed information of SVM model for classification between DS-TB and DR-TB

Environment	Intel Xeon CPU E3-1275 V6 @ 3.80 GHz 3.79 GHz	
Software	Matlab 2020a	
SVM model	Function	fitcsvm(trainData,trainlabel,'Standardize',true, ClassNames', {'sensitive', 'resistant'}, 'KernelFunction','RBF','KernelScale','auto');
	'Standardize'	Standardize the predictors before training using their corresponding weighted means and weighted standard deviations: $x_i^* = (x_i - \mu_i^*)/\sigma_i^*$, (1) where $\mu_i^* = \frac{1}{\sum_k w_k^*}\sum_k w_k^* x_{ik}$, (2) $(\sigma_i^*)^2 = \frac{v_1}{v_1^2 - v_2}\sum_k w_k^*(x_{ik} - \mu_i^*)^2$, (3) $v_1 = \sum_i w_i^*$, (4) $v_2 = \sum_i (w_i^*)^2$. (5) x_{ik} is observation k (row) of predictor i (column) and w_k^* is the weight of observation k (row). w_k^* is set to 1 in our work.
	Kernel function	RBF (Radial Basis Function) $h(x_i, x_j) = \exp(-\lVert x_i - x_j \rVert^2)$, (6)
	Optimization solver	SMO (Sequential Minimal Optimization)
	Class names	'sensitive'; 'resistant'
	No of predictors	25
	No of observations (1st round)	2618
	Weight matrix (1st round)	2618×1
	Support Vectors (1st round)	1934×25
	Support vector labels (1st round)	1934×1

Note: 25-feature-based training data and testing data for the 1st round of the ten-fold cross validation, as well as SVM testing code can be downloaded here: https://github.com/fyang5/ClassificationDRvsDS_SVM.git.