ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Fast Screening of Tuberculosis Patients Based on Analysis of Plasma by Infrared Spectroscopy Coupled with Machine Learning Approaches

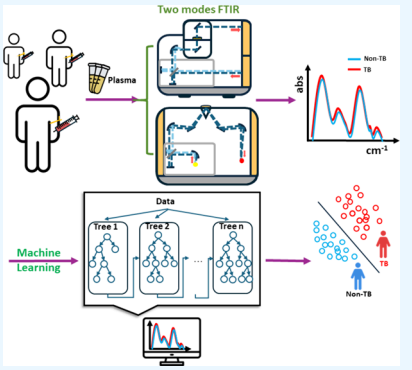
Mei Lin,◆ Hsiao-Chi Lu,◆ Hui-Wen Lin,* Sheng-Wei Pan,* Bing-Ming Cheng,* Ton-Rong Tseng, Jia-Yih Feng, and Mei-Lin Ho*

Cite This: *ACS Omega* 2025, 10, 11817−11827

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Prompt diagnosis of tuberculosis (TB) enables timely treatment, limiting spread and improving public health for this disease. Currently, a rapid, sensitive, accurate, and cost-effective detection of TB still remains a challenge. For this purpose, we engaged a transmission skill and an attenuated total reflectance (ATR) technique coupled with Fourier-transform infrared spectrometry (FTIR) to study the IR spectra of the plasma samples from TB patients (n = 10) and healthy individuals (n = 10). To ensure high-quality spectral data, spectra were collected in both transmission and ATR modes, with each measurement consisting of 256 scans at a resolution of 8 cm^{-1}. For the transmission mode, measurements were repeated five times per sample, while ATR-FTIR measurements were repeated three times per sample. These parameters were carefully optimized through rigorous testing to achieve the highest possible signal-to-noise ratio for patient sample analysis. Using this method, we obtained a total of 100 spectra from 20 samples in the transmission mode and 60 spectra in the ATR-FTIR mode, ensuring sufficient data for robust spectral analysis. Further, we applied machine learning techniques to analyze and classify the IR spectra; by this means, we differentiated those spectra between TB patients and healthy ones. In this work, we modified the transmission-FTIR setup to improve the absorption sensitivity by focusing the IR light on the interface of the sample; while, we used a high-refractive-index crystal ZnSe as a medium to reflect the signals in ATR scheme. Routinely, we compared the spectra obtained from both methods; in their second derivative curves, we notified that there had distinct spectral differences in protein and lipid regions (3500−3000, 2900−2800, and 1700−1500 cm^{-1}) between TB and healthy groups. Using three machine learning classifiers—Logistic Regression (LR), Random Forest (RF), and XGBoost (Xg)—we found that the Xg achieved an accuracy of 0.749, precision of 0.703, recall of 0.901, F1 score of 0.790, and an AUC of the ROC curve of 0.82 for absorption spectra in the 3500−2700 cm^{-1} region; additionally, the machine learning practice showed that ATR data possessed performance parameters of ∼ 80% in accuracy. We randomly assigned participants (rather than individual scans) to 80% training and 20% test sets to train and validate three machine learning models (LR, RF, and Xg). Based on the results, we concluded that the absorption spectroscopic method demonstrated its superior performance in TB diagnosis. Thus, we have showed that absorption-FTIR spectroscopy is a valuable tool for sorting the TB disease from patients. The spectral IR analysis of plasmas can complement clinical evidence and provides a rapid and accurate diagnosis of TB in clinic.



INTRODUCTION

Tuberculosis (TB) is a global epidemic caused by *Mycobacterium tuberculosis*, primarily infecting the lungs. According to the World Health Organization (WHO) Global TB Report 2023, the recounted number of individuals contracted with TB worldwide was 9.6 million in 2020, increasing to 10.3 million by 2021, and to 10.6 million by 2022.[1] To prevent the spread of this disease, one of a key factor is an early detection; then, to take an immediate treatment is possible and crucial. Current methods of detecting TB are based on chest X-ray radiography, sputum smear microscopy, mycobacterial culture, polymerase chain reaction, immunodiagnostic methods (tuberculin skin test and Interferon (IFN-γ) release assay),[2] and etc. Sputum acid-fast bacilli smear for pulmonary TB screening has a sensitivity of 10−75% and a specificity exceeding 90%, but its effectiveness declines in regions with high nontuberculous mycobacterium prevalence, necessitating rapid sputum TB PCR for confirmation.[3] Blood-based quantitative PCR can detect circulating *Mycobacterium tuberculosis* DNA, but its sensitivity, typically below 60%, limits its clinical utility.[4] To facilitate the diagnosis of

Received: August 30, 2024
Revised: March 11, 2025
Accepted: March 13, 2025
Published: March 20, 2025

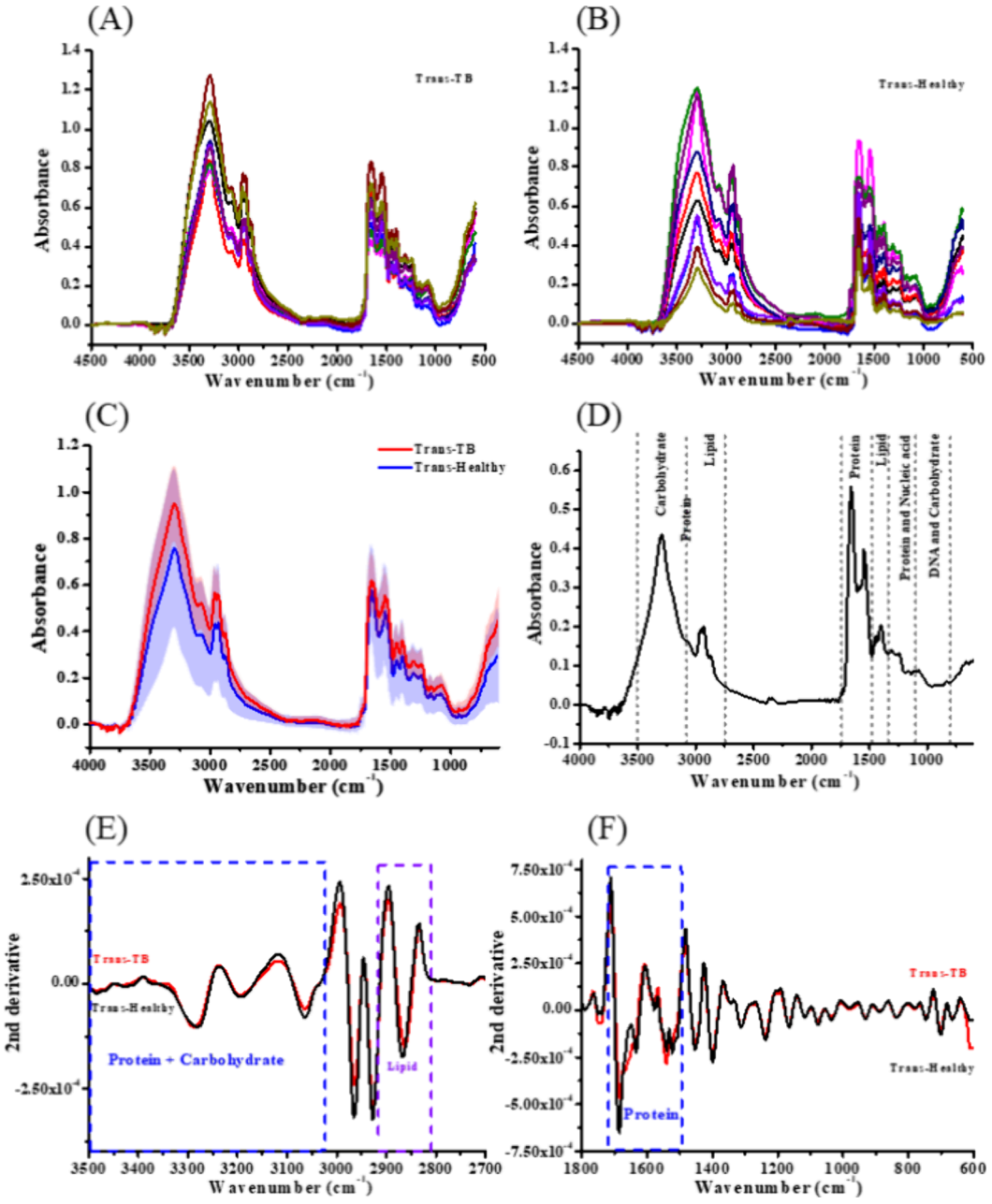


Figure 1. Direct Transmission-FTIR spectra of the dried films of plasma (A) from TB patients, (B) from healthy groups, (C) mean spectrum and its standard deviation for each group in (A) and (B), and (D) for the assignments of the main bands. The second derivatives of the averaged spectra in wavenumber region (E) 3500−2700 cm^{-1} and (F) 1800−600 cm^{-1}.

the TB, many new promising diagnostic techniques have been recently developed, including probe methods, CRISPR-Cas assay, mass spectrometry analysis, and immunosensor devices.[5] Nevertheless, a rapid, sensitive and simple diagnostic test that is applicable at point-of-care still needing development.

For the diagnosis of diseases, Fourier-transform infrared spectroscopy (FTIR) has been used as a spectral tool to analyze biological specimens for decades.[6−8] Various spectral techniques of FTIR provide the advantages of rapid identification with high sensitivity and repeatability; also, these transmission or reflectance methodologies of FTIR are nondestructive and require little or no preliminary preparation.[9,10] Usually, routine serological diagnosis offers a promising opportunity for increased testing accessibility, facilitated by simpler sample collection methods in hospitals; particularly, several studies have identified promising TB products in serum or plasma.[11] Recently, blood specimens have been studied with IR spectroscopy to develop better alternatives for disease diagnosis.[12] For example, IR spectroscopy has reportedly been used to analyze lung-related diseases.[13−15] Among these cases, Yang et al. proposed that the attenuated total reflectance of Fourier-transform infrared spectroscopy (ATR-FTIR) combined with chemometrics could be used effectively to distinguish the serum of patients with lung cancer from that of healthy people.[13] The findings indicated that the concentrations of protein, lipid and nucleic acid molecules in the serum of lung cancer patients were elevated in comparison to those of individuals without health issues. In addition, using infrared spectroscopy combined with chemometric analysis, Giamougiannis et al. studied blood plasma, blood serum, and urine samples from patients with

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

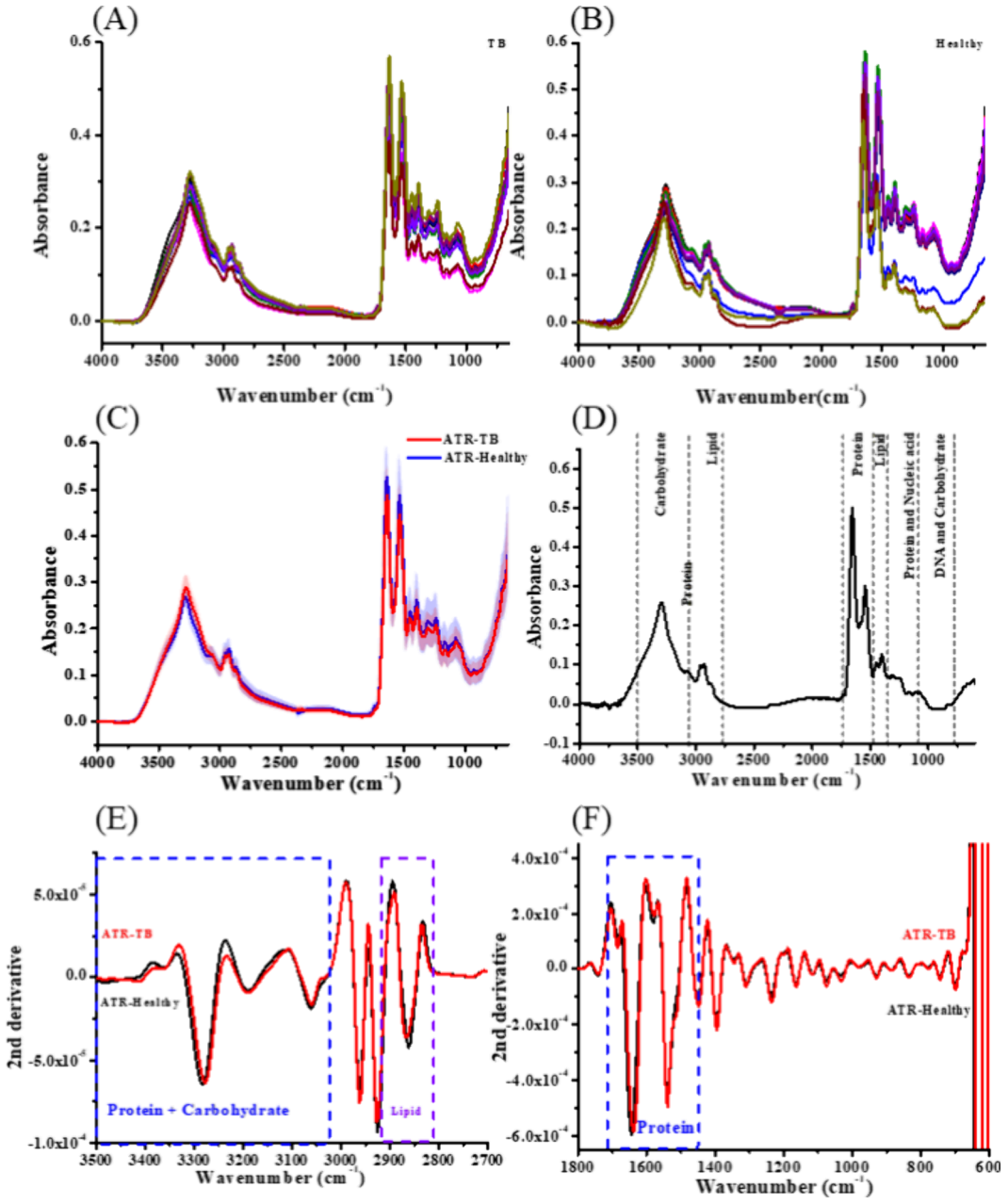


Figure 2. ATR-FTIR spectra of the dried films of plasma (A) from TB patients, (B) from healthy groups, (C) mean spectrum and its standard deviation for each group in (A) and (B), and (D) for the assignments of the main bands. The second derivatives of the averaged spectra in wavenumber region (E) 3500−2700 cm^{-1} and (F) 1800 cm^{-1}−600 cm^{-1}.

benign gynecological conditions and patients with ovarian cancer;[14] based on their findings, they supported the potential of ATR-FTIR spectroscopy to identify chemotherapy-related spectral changes. Adopting a different approach, Sitnikova et al.[15] investigated the ATR-FTIR spectra of serum samples with methods of multivariate processing and found differences in the ranges of DNA and RNA between the patients with breast cancer and healthy people. Based on their results, they concluded that the spectral technique could play a key role in breast cancer screening.

FTIR spectroscopy measures the absorption of infrared radiation for a molecule associated with its vibrations, such as stretching and bending, providing detailed insights into the chemical composition and structure of the molecule. The absorption band in the IR spectrum corresponds to specific functional groups, for examples, the O−H and C═O stretching bands reveal around 3200−3600 cm^{-1} and near 1700 cm^{-1}, respectively.[16] While FTIR offers high sensitivity and molecular specificity, challenges such as overlapping bands and matrix effects can complicate spectral interpretation.[17] To address these limitations, chemometric methods and machine learning algorithms, such as Principal Component Analysis (PCA), Random Forest (RF), and Support Vector Machines (SVM), are employed to extract meaningful patterns from complex spectral data.[18] These tools enhance the diagnostic potential of FTIR spectroscopy, making it a powerful approach for applications such as identifying disease-specific biomarkers and developing predictive diagnostic models.[19]

Integrating FTIR analysis with statistical methods, we can explore the relationships between spectra and biosample

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

constituents, thereby enhancing our understanding of the overall data results. Based on the existing literature, the combination of FTIR spectra with algorithms, chemometrics, or machine learning could be successfully applied to the detection, diagnosis, and discrimination of different diseases.[20−24] For instance, ATR-FTIR spectroscopy with a subsequent genetic algorithm−linear discriminant analysis algorithm was adopted to distinguish between the saliva swab samples of patients with COVID-19 and healthy controls.[20] Khoury et al. used RF classification in combination with FTIR spectra to detect multiple sclerosis and amyotrophic lateral sclerosis.[21] Gulekan et al. integrated FTIR spectra with a SVM model as a machine learning classification algorithm to examine blood serum and confirm primary myelofibrosis in patients.[22] The results from PCA indicated that it was possible to differentiate between patients and controls. Wu et al. performed serum analysis using FTIR and deep learning algorithms of the AlexNet, ResNet, MSCNN, and MSResNET models to differentiate among ankylosing spondylitis, rheumatoid arthritis, osteoarthritis, and healthy control groups successfully.[23] McHardy et al. added Wasserstein generative adversarial network-augmented ATR-FTIR spectra of each serum sample to improve the convolutional neural network of performance; by this means, they differentiated between pancreatic-cancer and noncancer samples.[24]

To the best of our knowledge, no published literature has addressed the blood analysis of TB patients using the FTIR method. This study investigates whether FTIR, paired with machine learning techniques, can be utilized as a rapid diagnostic tool to distinguish between plasma samples of TB patients and healthy controls. We employed the Logistic Regression (LR), Random Forest (RF) and XGBoost (Xg) models as machine learning classification algorithms to further analyze the absorption-FTIR and the ATR-FTIR spectral data, thereby demonstrating the potential of FTIR spectroscopy as a predictive tool for the diagnosis of TB.

RESULTS

Spectral Analysis. Figure 1A,B show the direct absorption spectra obtained from transmission mode FTIR of dried plasma films collected from TB patients and healthy individuals, respectively; Figure 1C depicts the mean spectrum and its standard deviation from the groups in Figure 1A,B. Similarly, Figures 2A−C show the ATR spectra of the ones from the TB patients, the healthy persons and mean spectrum and its standard deviation, respectively. For comparison, the intensities in Figure 1 are greater than those in Figure 2, possibly due to the optical lengths of the sample films in transmission mode being longer than the penetration depths of the films in the ATR method. Typically, the penetration depth of the film for the ATR-IR process is in the range of 2−5 μm, as it is not constant throughout the whole IR region. The penetration depth changes across the spectrum as a function of the wavelength of the IR light; the depth decreases with increasing wavenumber. Compared to the absorption spectrum, this effect results in diminished relative intensities of peaks at higher wavenumbers for the ATR-IR spectrum.

The main IR absorption bands all appeared in the averaged spectra of TB patients and healthy individuals taken in both modes; apparently, there were no significant differences in their absorption patterns, as can be seen from comparing Figures 1C and 2C. The typical main bands in each characteristic spectrum are marked in Figures 1D and 2D for the direct aabsorption and the ATR spectra, and the assignments of the mainbands are listed in Table 1.[25−27]

Table 1. Assignments of the Main Bands in the Relevant IR Absorption Spectra for Dried Plasma Films, Compiled Based on Data from References[25−27],[25−27a]

wavenumber	assignments	source describe
3294	ν(N−H)/ν(OH)	protein, carbohydrate
3078	ν(N−H)	protein
2958	$\nu_{as}(CH_3)$, $\nu_{as}(CH_2)$, ν_{as}(CH)	protein, lipid
2931	$\nu_{as}(CH_3)$, $\nu_{as}(CH_2)$, ν_{as}(CH)	protein, lipid
2873	$\nu_{sym}(CH_3)$, $\nu_{sym}(CH_2)$, ν_{sym}(CH)	protein, lipid
2858	$\nu_{sym}(CH_3)$, $\nu_{sym}(CH_2)$, ν_{sym}(CH)	protein, lipid
1735	ν(C═O)	cholesterol ester/triglyceride
1654	ν(C═O)	protein
1542	ν(CN), δ(NH)	protein
1454	$\delta(CH_2)$, $\delta(CH_3)$	lipid
1400	$\nu_{sym}(COO^-)$, $\delta(CH_3)$	protein, lipid
1342	$\delta(CH_2)$, $\delta(CH_3)$	lipid
1311	ν(CN), ν(N−H)	protein
1245	$\nu(PO_2^-)$	nucleic acid
1168	δ(COH), ν(CH), ν(CO), δ(COC)	carbohydrate, protein
1153	δ(COH), ν(CH), ν(CO), δ(COC)	carbohydrate
1103	δ(COH), ν(CH), ν(CO), δ(COC)	carbohydrate
1080	$\delta(PO_2^-)$, ν(CO), ν(COC), ν(COH)	phospholipid, carbohydrate, nucleic acid
1053	ν(CO), ν(COC), ν(COH)	carbohydrate
941	δ(COH), ν(CH), ν(CO), δ(COC)	DNA, carbohydrate

[a] δ: symmetric bending vibration; ν_{as}: asymmetric stretching vibration; ν_{sym}: symmetric stretching.

The absorption band at ∼3294/3078 cm^{-1} in Figure 1D may be associated with the vibrational modes such as N−H in protein amide groups and the stretching of −OH groups in carbohydrates. Attributed to the hydrogen bonding effect, this band is broad; also, the H-bond might cause a shift in the absorption band. The bands at 2958 and 2931 cm^{-1} are primarily attributed to the asymmetric stretching modes of CH groups in the structures of proteins and lipids, 2873 and 2858 cm^{-1} are symmetric stretching modes of CH groups in protein and lipid. Obviously, the band near 1735 cm^{-1} corresponds to the vibration of the C═O group originating from the substructures of cholesterol esters and triglycerides. In detail, the band around 1654 cm^{-1} is associated with protein amide groups; the one at 1542 cm^{-1} is related to stretching of C−N bonds and the bending vibration of N−H bonds in the amide II band; those at 1454 and 1342 cm^{-1} represent the absorptions from the constitution of the lipid; the one at 1400 cm^{-1} comes from the protein or lipid; and that at 1311 cm^{-1} is from the amide III related to the protein amide group. The bands at 1245, 1168, 1153, 1103, and 1080 cm^{-1} may mainly originate from nucleic acids, carbohydrates, proteins and phospholipids. In addition, the absorption at 1053 cm^{-1} is attributed to carbohydrates, and that at 941 cm^{-1} is from the DNA and carbohydrates.

We further examined the averaged spectra by analysis of the spectral curves for their second derivatives, which are partially presented in Figures 1E,F and 2E,F. In the second derivative

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

curves from the direct absorption in Figure 1F, when the curve from the TB patients was compared with that from the healthy individuals, there appeared to be significant spectral changes near the 1654 cm^{-1} region, which is responsive to the absorption of amides in proteins. Meanwhile, notable changes were observed in the 3500−3000 cm^{-1} region related to the absorption of the protein, as shown in Figure 1E. Also, the curves in the 2997−2965 cm^{-1} regions, corresponding to the vibrational asymmetric stretching of CH in lipids and 2900−2800 cm^{-1} regions are symmetric stretching modes of CH in protein and lipid. However, less change was noted in curves below the 1400 cm^{-1} regions, which are associated with nucleic acids, carbohydrates, and DNA.

In the case of the second derivative curves of the ATR-FTIR spectra, we observed patterns similar to that derived from the averaged spectrum of TB patients compared to the one from healthy individuals. Nevertheless, those curves revealed slight dissimilarity, as shown in Figure 2E,F.

Besides the variation in the spectral profiles, we also noted shifts of the lines for the second derivative curves in Figures 1E,F and 2E,F. For example, shifts of the lines occurred in the 1700−1500 and 3500−3000 cm^{-1} regions associated with the amide vibrational frequencies of proteins, and others occurred in the 2900−2800 cm^{-1} region due to the CH vibrations of lipids. These shifts of the lines of the second derivative curves derived from both absorption and ATR spectra indicate that the vibrational frequencies of protein structures and the CH vibrations of lipids might be affected by the intrinsic properties of the plasma in TB patients.

Integrated Report on Machine Learning Models for Spectral Data Analysis. In this study, we evaluated the performance of three machine learning models—LR, RF, and Xg—on analysis of absorption-FTIR and ATR-FTIR spectral data from 4500 to 600 cm^{-1} to distinguish the spectral variations in the plasma between TB patients and healthy individuals. The analytical results include a comprehensive comparison of performance metrics, ROC curves, and confusion matrices.

PERFORMANCE METRICS

For the absorption-FTIR data (Figure 3A), LR achieved an accuracy of 0.501, while RF and Xg showed higher accuracies at 0.706 and 0.726, respectively. Precision of LR was 0.502, compared to 0.650 for RF and 0.669 for Xg. Recall values were high across all models, with LR at 0.900 and both RF and Xg at 0.901. F1 scores were 0.645 for LR, 0.755 for RF, and 0.768 for Xg, indicating better overall performance for RF and Xg.

For the ATR-FTIR data (Figure 3B), LR achieved 0.527 in accuracy, while RF and Xg reached 0.661 and 0.697, respectively. Precision scores were 0.513 for LR, 0.606 for RF, and 0.638 for Xg. Recall values remained high, at 0.901 for LR, 0.903 for RF, and 0.901 for Xg. F1 scores were 0.654 for LR, 0.726 for RF, and 0.747 for Xg.

Comparisons of ROC Curves. ROC curves provided a visual representation of the models' abilities to distinguish between TB patients and healthy individuals. For the absorption-FTIR data (Figure 4A) the AUC was 0.57 for LR, while RF and Xg achieved higher values of 0.79 and 0.80, respectively. For the ATR-FTIR data (Figure 4B), the AUCs were 0.55 for LR, 0.76 for RF, and 0.79 for Xg. With the highest AUC, the Xg model exhibited superior generalization performance.

Confusion Matrices. Confusion matrices provide detailed insights into model classification performance. In the

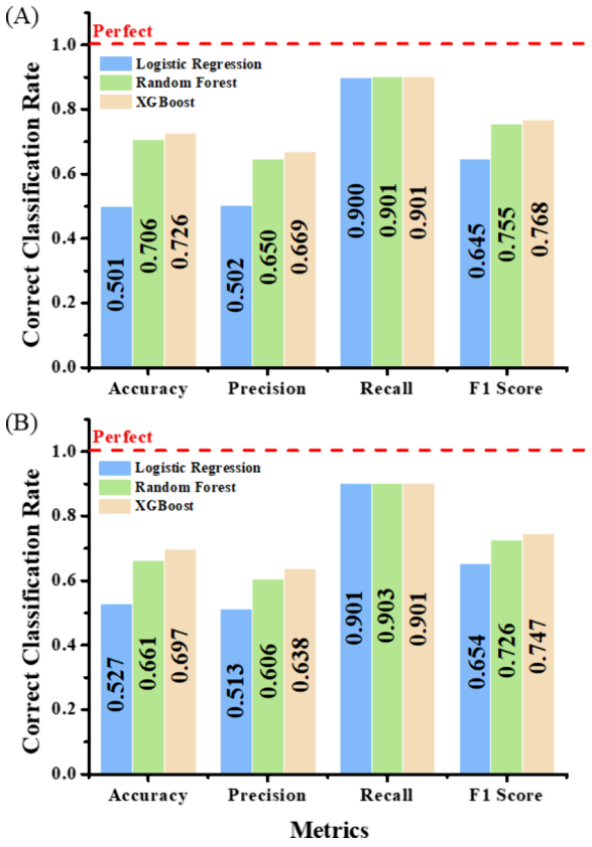


Figure 3. Comparison of accuracy, precision, recall, and F1 scores of (A) absorption-FTIR and (B) ATR-FTIR based on LR, RF, and Xg models.

absorption-FTIR testing set (Figure 5A), Xg achieved 1574 true positives and 946 true negatives, outperforming both LR and RF. In the training set, both RF and Xg demonstrated perfect classification with no errors, while LR showed substantial false positives and false negatives. For the ATR-FTIR testing set (Figure 5B), Xg again exhibited superior performance, with 1548 true positives and 869 true negatives. In the training set, RF and Xg maintained perfect classification, whereas LR continued to display numerous false positives and false negatives. Overall, Xg consistently outperformed the other models across both testing and training sets and for both types of spectral data.

Machine Learning in Different Spectral Regions. From the spectral analyses of the two methods, we further performed machine learning statistical analysis of plasma from TB patients and healthy individuals in the chosen region, and then we examined the performance indicators, AUCs of ROC curves, and the confusion matrix (Table 2, Figures S1 and S2). In the spectral region of 3500−2700 cm^{-1}, the performance of all three models' predictions was better than in the full region of 4500−600 cm^{-1}. Among the three models, Xg provided the best predictions. In the spectral region of 1800−600 cm^{-1}, the prediction performances of all three models were better than in the full region. Comparing the two regions of 3500−2700 and 1800−600 cm^{-1}, Xg also provided more accurate predictions in the former region, i.e., 3500−2700 cm^{-1}. The confusion matrices indicated that, except for the LR model, the RF and Xg models showed better classification performance in the 3500−2700 and 1800−600 cm^{-1} regions in the testing set, achieving 100% accuracy in the training set. On the other hand, further comparing ATR-FTIR analysis results (Table 2, Figures

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

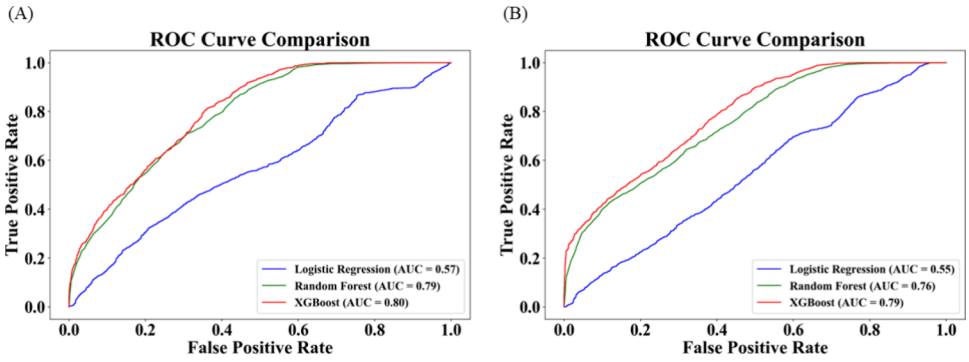


Figure 4. Plots the ROC curves of the models, and the AUC values represent the areas under the ROC curves. (A) absorption-FTIR and (B) ATR-FTIR.

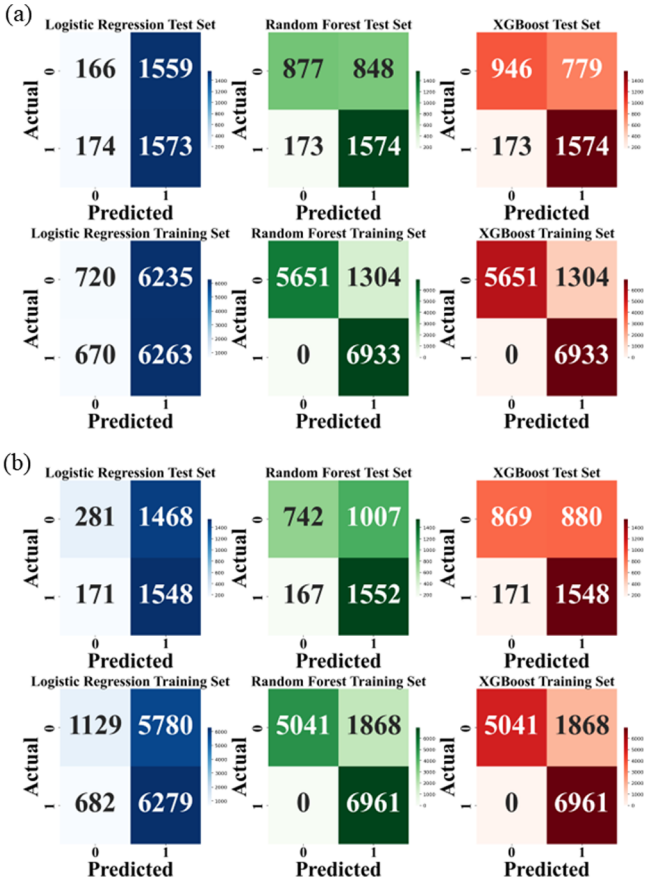


Figure 5. Confusion matrices of (a) absorption-FTIR and (b) ATR-FTIR for the models.

S3 and S4) revealed that the results in the confusion matrices were similar to those in the absorption spectra. The machine learning models performed better on the full region data than on the two segmented regions, with Xg making better predictions than the other two models.

DISCUSSION

Spectral Measurement. The FTIR spectra obtained from the plasma of TB patients and healthy individuals, whether in absorption-FTIR spectra or ATR mode, appear initially similar. This similarity is understandable, for the vibrations and rotations of various molecular groups overlap due to the complex components of whole blood.[28] However, the absorption intensity of each component is consistently higher in the plasma of patients with TB than in that of healthy individuals. This difference may be attributed to the FTIR spectra being strongly dominated by the high concentrations of proteins present in the serum or plasma, overshadowing other low molecular mass components.[29]

Further analysis using the second derivative processing of the spectra revealed structural changes in proteins (amide) and lipids in the plasma of TB patients compared to healthy individuals. This finding aligns with reports that patients with active TB have significantly higher levels of C-reactive protein (CRP),[30] and the associated IR spectral changes in CRP correlate with our observed spectral variations.[31] Furthermore, in our previous research,[4] we also found that in patients with TB, TB-specific interferon-γ, a cell signaling protein, is elevated, monocytes are increased, and lymphocytes are decreased, consistent with the IR results. However, it remains unclear whether CRP, interferon-γ, or both specific proteins are responsible for the spectral changes.

Additionally, regarding the lipid changes observed in the IR spectra, we hypothesize, based on our current knowledge, that they might be due to lipoarabinomannan (LAM). LAM is a glycolipid component of the cell wall of *Mycobacterium tuberculosis*, the bacteria that causes TB. Although it has been reported that LAM is excreted in urine in a soluble form and LAM assays can help diagnose TB,[32] we speculate that this component might also be present in plasma. In addition, the spectra measured from the plasma of healthy individuals were similar to those obtained from the standard plasma (SRM) purchased from NIST. This SRM represents "normal" human plasma and was obtained from 100 individuals. The SRM plasma samples from an equal number of men and women within a narrow adult age range (40−50 years) were used. The age distribution of our clinically healthy samples was similar to SRM samples, ensuring comparable spectral results. Although the TB group was significantly older than the control group, the absorption-FTIR (Figure 1A,B) and ATR-FTIR spectra (Figure 2A,B) in the wavenumber regions 3500−2700 and 1800−600 cm^{-1} showed no significant changes in peak trends. This indicates that age does not affect the spectral results.

Comparing the FTIR spectra from the two experimental modes revealed that the intensity is stronger in absorption-FTIR spectra than in the ATR mode. Although the intensity variations are more pronounced in direct absorption-FTIR spectra, the ATR mode provides more stable results.[16] Nevertheless, both modes yield similar conclusions, with the direct absorption-FTIR spectra mode displaying more distinct differences.

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

Table 2. Standard Performance Metrics Calculated for the Machine Learning Models

method	IR range (cm^{-1})	model	accuracy	precision	recall	F1 score	ROC AUC
absorption	4500−600	logistic regression	0.529	0.517	0.900	0.656	0.56
		random forest	0.705	0.647	0.903	0.754	0.78
		XGBoost	0.726	0.668	0.900	0.767	0.80
ATR	4500−600	logistic regression	0.527	0.513	0.901	0.654	0.55
		random forest	0.661	0.606	0.903	0.726	0.76
		XGBoost	0.697	0.638	0.901	0.747	0.79
absorption	3500−2700	logistic regression	0.681	0.639	0.901	0.748	0.66
		random forest	0.713	0.666	0.908	0.768	0.80
		XGBoost	0.749	0.703	0.901	0.790	0.82
ATR	3500−2700	logistic regression	0.587	0.566	0.901	0.695	0.61
		random forest	0.605	0.578	0.903	0.705	0.67
		XGBoost	0.625	0.592	0.901	0.715	0.72
absorption	1800−600	logistic regression	0.573	0.556	0.901	0.687	0.63
		random forest	0.729	0.681	0.903	0.776	0.79
		XGBoost	0.745	0.698	0.901	0.787	0.82
ATR	1800−750	logistic regression	0.502	0.508	0.901	0.649	0.54
		random forest	0.638	0.595	0.912	0.720	0.70
		XGBoost	0.615	0.580	0.901	0.706	0.70

Data Analysis. Recently, several teams have been actively working on integrating various experimental methods with advanced machine-learning algorithms to improve analysis accuracy.[33−35] In our study, to avoid overfitting risks, a 5-fold cross-validation approach was applied during model training, partitioning data into five subsets and ensuring each subset was used for validation. Additionally, ensemble methods like RF and Xg, known for reducing overfitting relative to single decision trees, were implemented to enhance generalization, particularly with limited data sets. Performance evaluation also included a 20% hold-out test set, ensuring unbiased assessment on unseen data. Beyond accuracy, metrics such as precision, recall, F1 score, and AUC-ROC were employed, with high recall indicating effective TB patient identification despite sample size limitations. And due to the limited number of cases, no drug-resistant TB cases were included in our study. A larger sample size study is warranted, incorporating different subtypes of TB patients, including pulmonary vs extrapulmonary TB and drug-resistant vs drug-sensitive cases, to provide more comprehensive insights.

To evaluate the differences between the models' performance, we employed the DeLong test[36] to compare their respective ROC curves. The results were as follows: LR vs RF yielded Z = 36.891 and a *p-value* = 0.0000, indicating RF significantly outperformed LR. Similarly, for LR vs Xg, Z = 36.528 and *p-value* = 0.0000 showed that Xg significantly outperformed LR. However, the comparison between RF and Xg resulted in Z = −0.332 and *p-value* = 0.7403, suggesting no significant difference between RF and Xg.

In addition to the DeLong test results, a comprehensive analysis of performance metrics, ROC curves, and confusion matrices demonstrated that Xg consistently outperformed LR and RF across both the absorption-FTIR and ATR-FTIR data sets. Specifically, the absorption-FTIR model exhibited higher predictive accuracy in the 2700−3500 cm^{-1} region, achieving an AUC of 0.82, while the ATR-FTIR model performed better across the full spectral range with an AUC of 0.79. These findings highlight the superior accuracy and robustness of the Xg model for spectral data analysis. The superior performance of Xg can be attributed to several factors. Its boosting mechanism iteratively refines the model by correcting errors from previous iterations, leading to higher accuracy and better AUC scores. Xg effectively captures complex, nonlinear patterns within the data and automatically identifies and leverages interactions between features, which is particularly crucial for spectral data. Additionally, Xg includes built-in regularization techniques that help prevent overfitting, ensuring that the model generalizes well to new data. Extensive options for hyperparameter tuning also allow for optimal model performance.

Moreover, the absorption-FTIR testing set demonstrated better machine learning analysis outcomes than those of the ATR-FTIR testing set. This can be attributed to the broader range of absorption intensity in absorption-FTIR, which provides richer spectral information. The enhanced spectral resolution allows machine learning models to identify subtle differences between TB patients and healthy individuals more effectively. The absorption-FTIR data, with its higher quality and detailed absorption features, enables models like Xg to leverage its advanced capabilities more fully, resulting in superior classification performance. In contrast, ATR-FTIR data, while useful, may not capture the same level of detail, leading to slightly lower performance metrics.

These advanced capabilities make Xg a robust and reliable model for distinguishing between TB patients and healthy individuals using spectral data, providing a valuable tool for medical diagnostics. The superior performance in absorption-FTIR analysis underscores the importance of high-resolution spectral data in improving the accuracy and reliability of machine learning models in medical diagnostics.

EXPERIMENTAL SECTIONS

Study Design and Enrollment. This case-control study included 10 patients with pulmonary TB and 8 healthy controls, all enrolled at the Department of Chest Medicine, Taipei Veterans General Hospital, Taiwan (ROC). The healthy controls had no active lung lesions, no history of TB diagnosis prior to enrollment, and remained free of TB during a 2-year follow-up. Additionally, two Standard Reference Material (SRM) 1950a human plasma samples from healthy individuals were obtained from NIST, bringing the total number of healthy subjects to 10.

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

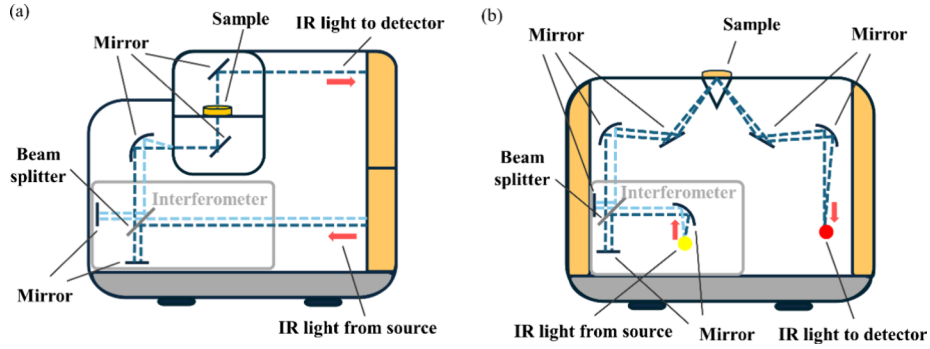


Figure 6. Optical layouts of FTIR systems for measurements of (a) an absorption spectrum in transmission mode, and (b) an attenuated total reflectance (ATR) spectrum.

Characteristics of Patients. The patients with TB were generally older than the controls (mean ± standard deviation, 63.3 ± 17.8 vs 47.9 ± 16.7 years, $p = 0.008$), but there were no statistical intergroup differences in the percentages of male gender (70% vs 50%, $p = 0.630$), smoking history (60% vs 13%, $p = 0.066$) or diabetes mellitus (20% vs 13%, $p = 1.000$), nor in BMI (20.8 ± 2.7 vs 26.4 ± 4.9 kg/m^2, $p = 0.078$). Furthermore, none of the participants had chronic kidney disease or chronic obstructive pulmonary disease.

Sample Preparation: Preparations of Plasma Samples. Ethical approval was granted by Taipei Veterans General Hospital (IRB No. 2021-05-007CC, 2021-05-006CC, 2022-06-011BC, and 2022-06-016CC); also, the consents were offered by all participants. The blood samples were kept in ethylenediaminetetraacetic acid (EDTA) tubes for plasma analysis. Within 2 h, the plasma samples were centrifuged at 1500 rpm for 10 min at 4 °C. The supernatants were stored at −80 °C until analysis.

Acquisitions of FTIR Spectra. Prior to analysis, the plasma samples were thawed in water bath for incubation at 25 °C for 30 min; then, the examinations were performed at room temperature.

Measurements of Absorption Spectra. The FTIR spectrometer was supplied by the ABB company (Quebec, Quebec, Canada) with a model 3000 system; which was further modified by Mastek Technologies, Inc. (XinZhuang, New Taipei City, Taiwan) for measurements of absorption spectra in a transmission mode. The optical layout of this system is shown in Figure 6A; in which, the FTIR was equipped with a liquid nitrogen-cooled MCT detector (Infrared Associates FTIR-16-1.0-LN2, Stuart, FL, USA). The plasma sample was applied onto a ZnSe substrate with a volume of 2 μL; then, the liquid sample was dried to a film by a fan for 10 min at least. To avoid interference from the environmental air, the FTIR and the sample chamber were purged with the high purity nitrogen gas. The spectrum was recorded in the wavenumber range of 4500−600 cm^{-1} with 256 scans at a resolution of 8 cm^{-1}. In addition, for the transmission mode, measurements were repeated five times per sample, while ATR-FTIR measurements were repeated three times per sample, and averaged by the software Origin Pro 8.5 to derive the final spectrum for each sample.

Measurements of ATR Spectra. The ATR-IR spectra of plasma samples were obtained with the FTIR (model FTLA2000-104) provided by the ABB company; in which, the MCT was also used as the detector. The optical layout of the ATR-FTIR system is illustrated in Figure 6B. A single-bunch ATR accessory with a ZnSe crystal (Pike Technologies, Madison, WI, USA) was attached to the FTIR. For the sampling, the 2 μL liquid plasma sample was dropped onto the top surface of the ZnSe crystal by a microsyringe; subsequently, the liquid sample was also dried as the film with the fan for 5 min at least. Each ATR spectrum was also determined with 256 scans at a resolution of 8 cm^{-1} in the wavelength range of 4500−600 cm^{-1}. For the transmission mode, measurements were repeated five times per sample, while ATR-FTIR measurements were repeated three times per sample. After the measurement of each sample, the crystal was cleaned first with optical grade acetone and then with optical grade ethanol several times to avoid intersample contamination.

Data Analysis. The spectral data were processed with baseline correction and the mean of quintuplicate; in addition, the second derivative curve was preprocessed with a 10-point Savitzky−Golay polynomial to separate overlapping bands and increase the apparent spectral resolution.

Modeling and Testing Procedure. We recruit participants (TB patients and healthy controls) and measure their plasma samples using FTIR at multiple wavenumber positions $\{\omega_1, \omega_2, \cdots, \omega_d\}$. Each participant i thus yields a set of intensities $\{x_{i1}, x_{i2}, \cdots, x_{id}\}$. Depending on our chosen approach, we might treat each wavenumber reading as a separate data point or compile all wavenumbers into a single vector. Regardless, the crucial point is that each participant's data remain grouped together for train/test partitioning.

1. Train/Test Partition
   (1) We randomly assign each participant i to either the training set $\mathcal{I}_{\text{train}}$ or the test set $\mathcal{I}_{\text{test}}$.
   (2) All wavenumbers from participant i go to the same set—none of participant i's data are split across train and test. This step avoids data leakage and the so-called "replicate sample trap."
2. Model Building Using the Training Set
   (1) Only participants in $\mathcal{I}_{\text{train}}$ are used for both model calibration and hyperparameter optimization.
   (2) Depending on the model (e.g., Logistic Regression, Random Forest, XGBoost), we follow these substeps:
      i. Feature Scaling: We may apply standardization to the training data wavenumbers/intensities.
      ii. Hyperparameter Tuning:
      We use cross-validation, testing various parameters (e.g., learning rate, regularization).

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

iii. Final Model Selection: After deciding the best hyperparameters, we retrain the model on the entire training set to finalize the parameters.

3. Testing (Evaluation) Using the Held-Out Test Set
   (1) The final model is then applied to the participants in $\mathcal{I}_{test}$.
   (2) We compute performance metrics:
      i. Accuracy, Precision, Recall (Sensitivity), F1
      ii. Confusion Matrix to visualize true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN).

ROC Curve & AUC to assess the trade-off between True Positive Rate (Recall) and False Positive Rate.

Classification and Validation Models. The FTIR spectra from the patients and healthy individuals were analyzed in the Python programming language. The spectra were investigated with three machine learning models: Random Forest (RF), Logistic Regression (LR), and Extreme gradient boosting (Xg).

Logistic Regression (LR). The LR model, a common approach in probabilistic nonlinear regression, is frequently used for analyzing and classifying binary and proportional response data sets. Here, the LR model was used for investigation of the relationship between binominal observation results (wavenumber and absorption value in both sets of IR spectra.

Random Forest (RF). The RF classifier is an ensemble learning method known for its robustness and ability to handle complex data sets. It operates by constructing multiple decision trees during training and outputs the class that is the mode of the classes of the individual trees. This approach helps in reducing overfitting and improving model accuracy. According to the existing literature, the RF algorithm has been applied in the discrimination of multiple sclerosis and amyotrophic lateral sclerosis[21] and in the diagnosis of transformation stages in esophageal squamous cell carcinoma tissue.[37] Therefore, the model was chosen to differentiate the plasma samples of patients and healthy individuals.

The important operation parameters of RF are summarized below:

1. Number of Trees: The number of trees in the forest was a key parameter. A higher number of trees generally improves the model's performance but also increases the computational complexity. For this study, we used the default setting provided by the sklearn library, which is a typical balanced choice for many data sets.
2. Maximum Depth of Trees: This parameter determines the maximum depth of each tree. Allowing the trees to grow too deep can lead to overfitting. We used the default setting, which allows the trees to expand until all leaves contain less than the minimum samples required, to split a node.
3. Bootstrap Sample: We used the default bootstrap sampling method, wherein each tree in the forest is built from a bootstrap sample from the data. This means that some data points will be sampled multiple times for one tree, while others may not by sampled at all, helping to increase diversity among the trees.

Extreme Gradient Boosting (Xg). The Xg algorithm works by dividing the original data set into multiple subdata sets. Each subdata set is randomly assigned to the classification and regression tree (CART) for predication. The results are calculated by the CART according to a certain weight and then are summed up as the final predictive output.

For the model analysis, the data were separated into a training set (80%) and a testing set (20%) by the Kennard−Stone method (Table 3).

Table 3. Division of Training Set and Testing Set

sample	total sample	TB	healthy
total sample	17,360	8680	8680
training set	13,888	6933	6955
testing set	3472	1747	1725

Evaluation Index of Model Parameters. The accuracy, precision, specificity, sensitivity, confusion matrix, recall, F1 score, receiver operating characteristic (ROC) curve, and area under the ROC curve (AUC) are important evaluation indicators for accessing the performance of machine learning models. The confusion matrix (Figure 7) was used to evaluate the classification models. A confusion matrix shows how well a model performs. Accuracy shows how often a classification model is correct overall. Precision shows how often predictions for the positive class are correct. Sensitivity, also known as recall or true positive rate, is how well the model finds all positive instances in the data set. Specificity is the probability that a test will be negative when the disease is not present. An F1 score is a harmonic mean of the precision and sensitivity. The best value is 1 and the worst is 0. Larger values of the parameters indicate better diagnostic performance of the model. The formulas for each parameter are shown in eqs 1−5. The ROC curve is the plot of the model's true positive rate (sensitivity) against its false positive rate (1-specificity) at each threshold setting, representing the effectiveness of the binary classification model. AUC reflects the scales with the overall classification performance. The AUC curve is 1.0 for a perfect model and 0.5 for a guessing model. A greater AUC value denotes better model performance.

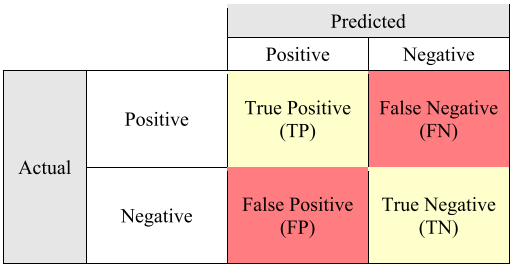


Figure 7. Confusion matrix. TP = True positive, FN = False negative, TN = True negative, and FP = False positive.

$$\text{Accuracy} = \frac{n_{\text{correct}}}{n_{\text{total}}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

n is the number of predictions

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Sensitivity}\ (\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \quad (3)$$

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827

$$\text{Specificity}\ (\%) = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \quad (4)$$

$$F1 = 2\frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (5)$$

CONCLUSIONS

This study integrates two FTIR methods (direct absorption-FTIR and ATR spectra) with machine learning techniques to develop models for classifying and predicting plasma from TB patients and healthy individuals. Both IR methods reveal differences in the structures of proteins and lipids between patients with TB and healthy individuals. Statistical analysis demonstrates that combining IR methods with machine learning can be used effectively to classify plasma samples from TB patients and healthy individuals, with the transmission mode outperforming ATR-FTIR. With the use of Xg in the protein expression region of 3500−2700 cm^{-1}, the performance indicators, specifically the accuracy of TB prediction, achieve an AUC score of 0.80. This study confirms that Xg is a valuable tool for distinguishing between TB patients and healthy individuals. Therefore, the combination of absorption IR and the Xg model offers a rapid method of TB detection.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c07990.

Figures S1 and S2: Performance indicators (A), ROC curve (B), and confusion matrix (C) of absorption-FTIR in the range of 3500−2700 and 1800−600 cm^{-1} for the models. Figures S3 and S4: Performance indicators (A), ROC curve (B), and confusion matrix (C) of ATR-FTIR in the range of 3500−2700 and 1800−750 cm^{-1} for the models (PDF)

AUTHOR INFORMATION

Corresponding Authors

Hui-Wen Lin − *Department of Mathematics, Soochow University, Taipei 111, Taiwan*; Email: hwlin@scu.edu.tw

Sheng-Wei Pan − *Department of Chest Medicine, Taipei Veterans General Hospital, Taipei 11217, Taiwan; School of Medicine, National Yang Ming Chiao Tung University, Taipei 12304, Taiwan*; Email: swpan2@vghtpe.gov.tw

Bing-Ming Cheng − *Department of Medical Research, Hualien Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Hualien City 97002, Taiwan; Center for General Education, Tzu Chi University, Hualien City 97005, Taiwan*; orcid.org/0000-0002-8540-6274; Email: bmcheng7323@gmail.com

Mei-Lin Ho − *Department of Chemistry, Fu Jen Catholic University, New Taipei City 242, Taiwan*; orcid.org/0009-0006-9415-698X; Email: 157382@mail.fju.edu.tw

Authors

Mei Lin − *Department of Chemistry, Fu Jen Catholic University, New Taipei City 242, Taiwan*

Hsiao-Chi Lu − *Department of Medical Research, Hualien Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Hualien City 97002, Taiwan*; orcid.org/0000-0002-0280-1317

Ton-Rong Tseng − *Mastek Technologies, Inc., New Taipei City 24892, Taiwan*

Jia-Yih Feng − *Department of Chest Medicine, Taipei Veterans General Hospital, Taipei 11217, Taiwan; School of Medicine, National Yang Ming Chiao Tung University, Taipei 12304, Taiwan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c07990

Author Contributions

◆M.L. and H.-C.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The research was supported by the National Science and Technology Council, Taiwan.

REFERENCES

(1) Huang, H.; Chen, Y.; Zuo, J.; Deng, C.; Fan, J.; Bai, L.; Guo, S. MXene-incorporated C60NPs and Au@Pt with dual-electric signal outputs for accurate detection of Mycobacterium tuberculosis ESAT-6 antigen. *Biosens. Bioelectron.* **2023**, *242*, No. 115734.

(2) Saminathan, S.; Panse, D.; Krishnappa, P. Current Trend in Diagnosis of Tuberculosis Infection. *Int. J. Collab. Res. Int. Med. Public Health* **2019**, *11*, 898.

(3) Chen, W.-C.; Chang, C.-C.; Lin, Y. E. Pulmonary Tuberculosis Diagnosis Using an Intelligent Microscopy Scanner and Image Recognition Model for Improved Acid-Fast Bacilli Detection in Smears. *Microorganisms* **2024**, *12* (8), 1734.

(4) Pan, S.-W.; Su, W.-J.; Chan, Y.-J.; Chuang, F.-Y.; Feng, J.-Y.; Chen, Y.-M. Mycobacterium tuberculosis−derived circulating cell-free DNA in patients with pulmonary tuberculosis and persons with latent tuberculosis infection. *PLoS One* **2021**, *16* (6), No. e0253879.

(5) Dong, B.; He, Z.; Li, Y.; Xu, X.; Wang, C.; Zeng, J. Improved Conventional and New Approaches in the Diagnosis of Tuberculosis. *Front. Microbiol.* **2022**, *13*, No. 924410.

(6) Woernley, D. L. Infrared absorption curves for normal and neoplastic tissues and related biological substances. *Cancer Res.* **1952**, *12* (7), 516.

(7) Nogueira, M. S.; Barreto, A. L.; Furukawa, M.; Rovai, E. S.; Bastos, A.; Bertoncello, G.; Carvalho, L. FTIR spectroscopy as a point of care diagnostic tool for diabetes and periodontitis: A saliva analysis approach *Photodiagn. Photodyn. Ther.* **2022**, *40*, No. 103036.

(8) Paraskevaidi, M.; Karim, S.; Santos, M.; Lima, K.; Crean, S. The use of ATR-FTIR spectroscopy for the diagnosis of Alzheimer's disease using oral buccal cells. *Appl. Spectrosc. Rev.* **2024**, *59* (8), 1021.

(9) Chen, H.; Li, X.; Zhang, S.; Yang, H.; Gao, Q.; Zhou, F. Rapid and sensitive detection of esophageal cancer by FTIR spectroscopy of serum and plasma. *Photodiagn. Photodyn. Ther.* **2022**, *40*, No. 103177.

(10) Sala, A.; Anderson, D. J.; Brennan, P. M.; Butler, H. J.; Cameron, J. M.; Jenkinson, M. D.; Rinaldi, C.; Theakstone, A. G.; Baker, M. J. Biofluid diagnostics by FTIR spectroscopy: A platform technology for cancer detection. *Cancer Lett.* **2020**, *477*, 122.

(11) Chendi, B. H.; Snyders, C. I.; Tonby, K.; Jenum, S.; Kidd, M.; Walzl, G.; Chegou, N. N.; Dyrhol-Riise, A. M. A Plasma 5-Marker Host Biosignature Identifies Tuberculosis in High and Low Endemic Countries. *Front. Immunol.* **2021**, *12*, No. 608846.

(12) Su, K.-Y.; Lee, W.-L. Fourier Transform Infrared Spectroscopy as a Cancer Screening and Diagnostic Tool: A Review and Prospects. *Cancers* **2020**, *12* (1), 115.

(13) Yang, X.; Qu, Q.; Qian, K.; Yang, J.; Bai, Z.; Yang, W.; Shi, Y.; Liu, G. Diagnosis of Lung Cancer by ATR-FTIR Spectroscopy and Chemometrics. *Front. Oncol.* **2021**, *11*, No. 753791.

(14) Giamougiannis, P.; Morais, C. L. M.; Rodriguez, B.; Wood, N. J.; Martin-Hirsch, P. L.; Martin, F. L. Detection of ovarian cancer (± neo-adjuvant chemotherapy effects) via ATR-FTIR spectroscopy: comparative analysis of blood and urine biofluids in a large patient cohort. *Anal. Bioanal. Chem.* **2021**, *413*, 5095.

(15) Sitnikova, V. E.; Kotkova, M. A.; Nosenko, T. N.; Kotkova, T. N.; Martynova, D. M.; Uspenskaya, M. V. Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis. *Talanta* **2020**, *214*, No. 120857.
(16) Pan, S.-W.; Lu, H.-C.; Lo, J.-I.; Ho, L.-I.; Tseng, T.-R.; Ho, M.-L.; Cheng, B.-M. Using an ATR-FTIR Technique to Detect Pathogens in Patients with Urinary Tract Infections: A Pilot Study. *Sensors* **2022**, *22* (10), 3638.
(17) Liu, X.; Renard, C.; Bureau, S.; Le Bourvellec, C. Revisiting the contribution of ATR-FTIR spectroscopy to characterize plant cell wall polysaccharides. *Carbohydr. Polym.* **2021**, *262*, No. 117935.
(18) Du, Y.; Hua, Z.; Liu, C.; Lv, R.; Jia, W.; Su, M. ATR-FTIR combined with machine learning for the fast non-targeted screening of new psychoactive substances. *Forensic Sci. Int.* **2023**, *349*, No. 111761.
(19) Singh, A.; Pandey, B. An Efficient Diagnosis System for Detection of Liver Disease Using a Novel Integrated Method Based on Principal Component Analysis and K-Nearest Neighbor (PCA-KNN). *International Journal of Healthcare Information Systems and Informatics* **2016**, *11* (4), 56.
(20) Barauna, V. G.; Singh, M. N.; Barbosa, L. L.; Marcarini, W. D.; Vassallo, P. F.; Mill, J. G.; Ribeiro-Rodrigues, R.; Campos, L. C. G.; Warnke, P. H.; Martin, F. L. Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity. *Anal. Chem.* **2021**, *93* (5), 2950.
(21) Khoury, Y. E.; Collongues, N.; Sèze, J. D.; Gulsari, V.; Patte-Mensah, C.; Marcou, G.; Varnek, A.; Mensah-Nyagan, A. G.; Hellwig, P. Serum-based differentiation between multiple sclerosis and amyotrophic lateral sclerosis by Random Forest classification of FTIR spectra. *Analyst* **2019**, *144* (15), 4647.
(22) Guleken, Z.; Ceylan, Z.; Aday, A.; Bayrak, A. G.; Hindilerden, İ. Y.; Nalçacı, M.; Jakubczyk, P.; Jakubczyk, D.; Depciuch, J. Application of Fourier Transform InfraRed spectroscopy of machine learning with Support Vector Machine and principal components analysis to detect biochemical changes in dried serum of patients with primary myelofibrosis. *Biochim. Biophys. Acta* **2023**, *1867* (10), No. 130438.
(23) Wu, X.; Shuai, W.; Chen, C.; Chen, X.; Luo, C.; Chen, Y.; Shi, Y.; Li, Z.; Lv, X.; Chen, C.; et al. Rapid screening for autoimmune diseases using Fourier transform infrared spectroscopy and deep learning algorithms. *Front. Immunol.* **2023**, *14*, No. 1328228.
(24) McHardy, R. G.; Antoniou, G.; Conn, J. J. A.; Baker, M. J.; Palmer, D. S. Augmentation of FTIR spectral datasets using Wasserstein generative adversarial networks for cancer liquid biopsies. *Analyst* **2023**, *148* (16), 3860.
(25) Cortizas, A. M.; López-Costas, O. Linking structural and compositional changes in archaeological human bone collagen: an FTIR-ATR approach. *Sci. Rep.* **2020**, *10*, 17888.
(26) Li, H.; Wang, J.; Li, X.; Zhu, X.; Guo, S.; Wang, H.; Yu, J.; Ye, X.; He, F. Comparison of serum from lung cancer patients and from patients with benign lung nodule using FTIR spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2024**, *306*, No. 123596.
(27) Araújo, R.; Ramalhete, L.; Ribeiro, E.; Calado, C. Plasma versus Serum Analysis by FTIR Spectroscopy to Capture the Human Physiological State. *BioTech* **2022**, *11* (4), 56.
(28) Guang, P.; Huang, W.; Guo, L.; Yang, X.; Huang, F.; Yang, M.; Wen, W.; Li, L. Blood-based FTIR-ATR spectroscopy coupled with extreme gradient boosting for the diagnosis of type 2 diabetes. *Medicine* **2020**, *99* (15), No. e19657.
(29) Bel'skaya, L. V.; Sarf, E. A.; Solomatin, D. V. Application of FTIR Spectroscopy for Quantitative Analysis of Blood Serum: A Preliminary Study. *Diagnostics* **2021**, *11* (12), 2391.
(30) Mousavian, Z.; Folkesson, E.; Fröberg, G.; Foroogh, F.; Correia-Neves, M.; Bruchfeld, J.; Källenius, G.; Sundling, C. A protein signature associated with active tuberculosis identified by plasma profiling and network-based analysis. *iScience* **2022**, *25* (12), No. 105652.
(31) Mateescu, A. L.; Mincu, N.-B.; Vasilca, S.; Apetrei, R.; Stan, D.; Zorilă, B.; Stan, D. The influence of sugar-protein complexes on the thermostability of C-reactive protein (CRP). *Sci. Rep.* **2021**, *11*, 13017.
(32) Flores, J.; Cancino, J. C.; Chavez-Galan, L. Lipoarabinomannan as a Point-of-Care Assay for Diagnosis of Tuberculosis: How Far Are We to Use It? *Front. Microbiol.* **2021**, *12*, No. 638047.
(33) Liu, Y.; Wang, R.; Zhang, C.; Huang, L.; Chen, J.; Zeng, Y.; Chen, H.; Wang, G.; Qian, K.; Huang, P. Automated Diagnosis and Phenotyping of Tuberculosis Using Serum Metabolic Fingerprints. *Adv. Sci.* **2024**, *11* (39), No. 2406233.
(34) Zhu, Y.; Girault, H. H. Algorithms push forward the application of MALDI−TOF mass fingerprinting in rapid precise diagnosis. *VIEW* **2023**, *4* (2), No. 20220042.
(35) Chen, X.; Shu, W.; Zhao, L.; Wan, J. Advanced mass spectrometric and spectroscopic methods coupled with machine learning for in vitro diagnosis. *VIEW* **2023**, *4* (1), No. 20220038.
(36) DeLong, E. R.; DeLong, D. M.; Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **1988**, *44* (3), 837.
(37) Yang, H.; Li, X.; Zhang, S.; Li, Y.; Zhu, Z.; Shen, J.; Dai, N.; Zhou, F. A one-dimensional convolutional neural network based deep learning for high accuracy classification of transformation stages in esophageal squamous cell carcinoma tissue using micro-FTIR *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2023**, *289*, No. 122210.

https://doi.org/10.1021/acsomega.4c07990
ACS Omega 2025, 10, 11817−11827