



OPEN ACCESS

EDITED BY

Simone Furini,
University of Bologna, Italy

REVIEWED BY

Mukul Midha,
Institute for Systems Biology (ISB),
United States
Deepak Ranjan Nayak,
Indian Institute of Information
Technology Design and Manufacturing
Kancheepuram, India

*CORRESPONDENCE

Jie Liu,
✉ 1790216088@qq.com

SPECIALTY SECTION

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 09 November 2022

ACCEPTED 27 February 2023

PUBLISHED 09 March 2023

CITATION

Zhu Q and Liu J (2023), A united model
for diagnosing pulmonary tuberculosis
with random forest and artificial
neural network.
Front. Genet. 14:1094099.
doi: 10.3389/fgene.2023.1094099

COPYRIGHT

© 2023 Zhu and Liu. This is an open-
access article distributed under the terms
of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A united model for diagnosing pulmonary tuberculosis with random forest and artificial neural network

Qingqing Zhu and Jie Liu*

Anhui Provincial Tuberculosis Institute, Hefei, Anhui, China

Background: Pulmonary tuberculosis (PTB) is a chronic infectious disease and is the most common type of TB. Although the sputum smear test is a gold standard for diagnosing PTB, the method has numerous limitations, including low sensitivity, low specificity, and insufficient samples.

Methods: The present study aimed to identify specific biomarkers of PTB and construct a model for diagnosing PTB by combining random forest (RF) and artificial neural network (ANN) algorithms. Two publicly available cohorts of TB, namely, the GSE83456 (training) and GSE42834 (validation) cohorts, were retrieved from the Gene Expression Omnibus (GEO) database. A total of 45 and 61 differentially expressed genes (DEGs) were identified between the PTB and control samples, respectively, by screening the GSE83456 cohort. An RF classifier was used for identifying specific biomarkers, following which an ANN-based classification model was constructed for identifying PTB samples. The accuracy of the ANN model was validated using the receiver operating characteristic (ROC) curve. The proportion of 22 types of immunocytes in the PTB samples was measured using the CIBERSORT algorithm, and the correlations between the immunocytes were determined.

Results: Differential analysis revealed that 11 and 22 DEGs were upregulated and downregulated, respectively, and 11 biomarkers specific to PTB were identified by the RF classifier. The weights of these biomarkers were determined and an ANN-based classification model was subsequently constructed. The model exhibited outstanding performance, as revealed by the area under the curve (AUC), which was 1.000 for the training cohort. The AUC of the validation cohort was 0.946, which further confirmed the accuracy of the model.

Conclusion: Altogether, the present study successfully identified specific genetic biomarkers of PTB and constructed a highly accurate model for the diagnosis of PTB based on blood samples. The model developed herein can serve as a reliable reference for the early detection of PTB and provide novel perspectives into the pathogenesis of PTB.

KEYWORDS

pulmonary tuberculosis, diagnosis, biomarker, RF, ANN

Introduction

Tuberculosis (TB) affects nearly five million adult males, 3.5 million adult females, and 1 million children, and there are approximately 10.4 million cases of TB worldwide (Jeremiah et al., 2022). Owing to the increasing global population, public health departments are continually aiming to improve the diagnostic efficiency of TB and reduce its rate of transmission. Microscopic examination of sputum smears for acid-fast bacilli and sputum cultures are commonly used for diagnosing pulmonary TB (PTB) worldwide. However, these microbiology-based approaches and culture methods are time consuming, and the probability of infection is high (Barac et al., 2019). It is therefore urgently necessary to study and development non-sputum-based, simple, sensitive, and specific tests for diagnosing PTB. The biomarkers of PTB have been increasingly explored in the last three years owing to several studies on the identification of novel diagnostic biomarkers and development of novel diagnostic methods for PTB (Khambati et al., 2021; Morrison and Mcshane, 2021; Khimova et al., 2022). These studies have paved the way for the diagnosis and identification of novel biomarkers of PTB. While there has been success in clinical use of pathogen-based biomarkers in the form of Cepheid GeneXpert and Urine Lipoarabinomannan (LAM), host-based biomarkers are in less advanced stages of development (Nogueira et al., 2022). Based on previous literature, the present study aimed to identify more specific biomarkers of PTB using blood samples.

Blood-based gene expression signatures are the most potential biomarkers for diagnosing PTB. According to the target product profile (TPP) for non-sputum biomarker triage tests published by the World Health Organization in April 2014, TPPs require a minimum diagnostic sensitivity of 90% and specificity of 87% for the diagnosis of PTB in adults (Denkinger et al., 2019). Several recent studies have demonstrated that whole-blood RNA signatures can be used for predicting active TB infections and determining the progression of *Mycobacterium tuberculosis* infections in individuals who are at a risk of developing active TB (Kaforou et al., 2013; Blankley et al., 2016; Sweeney et al., 2016; Zak et al., 2016).

The increasing use of high-throughput sequencing technologies in the last decade has enabled the investigation of various aspects of diverse diseases (Dillies et al., 2013; Sullivan et al., 2017). Large volumes of high-throughput data have been stored in public platforms owing to the rapid development of high-throughput sequencing technology. These data can therefore be used for selecting critical indicators or feature biomarkers, which is a significant challenge for the development of diagnostic models. Machine learning techniques, including random forest (RF) and artificial neural network (ANN), can provide novel insights for solving this problem, and have been widely employed in previous studies for constructing diagnostic models by analyzing sequencing data (Dillies et al., 2013; Sullivan et al., 2017). Random Forest algorithm can perform random sampling to screen the target biomarkers and has high predicted accuracy (Byeon, 2019). Furthermore, the Artificial Neural Network can be used to evaluate the weight of target biomarkers screened by RF and

construct the predicted model for PTB with divided training and validation datasets (Curchoe et al., 2020). However, multi-biomarker-based diagnostic models and the combination of RF and ANN have not been employed for the diagnosis of TB to date.

The present study aimed to construct a multi-mRNA diagnostic model for the diagnosis of PTB. To this end, the genes that were differentially expressed between the PTB and control samples in the public datasets in the Gene Expression Omnibus (GEO) database were initially identified. The essential biomarkers for classifying PTB were screened using an RF classifier, and the weight of each biomarker was determined using ANN. A diagnostic model was subsequently developed based on these biomarkers and the accuracy of the model in discriminating between PTB and control samples was verified by receiver operating characteristic (ROC) curve analysis. The area under the curve (AUC) of the training (GSE83456) and validation (GSE42834) cohorts was determined to be 1.000 and 0.946, respectively. The high accuracy indicated that the diagnostic model constructed herein met the necessary requirements for the clinical diagnosis of PTB. The protocol and algorithms used in the present study are depicted in Figure 1.

Methods

Data processing

In this study, two RNA expression datasets were initially retrieved from the GEO database using the keywords “tuberculosis, normal.” The GSE83456 and GSE42834 datasets were processed using the GPL10558 platform of an Illumina HumanHT-12 V4.0 Expression BeadChip system. Based on the available literature on the use of machine learning for the diagnosis of diseases, we assumed that the sample size of the two datasets was appropriate for developing a machine learning-based diagnostic model. The obtained RNA-seq data were subsequently annotated and normalized using R software (version 4.2.1). The GSE83456 and GSE42834 datasets were selected as the training and validation cohorts, respectively.

Identification of differentially expressed genes (DEGs)

The DEGs between the PTB and control samples in the training set were identified using the limma package in R, with $p < 0.05$ and $|\log_2\text{foldchange (FC)}| > 1.0$. The DEGs were visualized using the pheatmap and ggplot2 packages in R.

Functional enrichment analysis

The identified DEGs were subjected to Gene Ontology (GO) enrichment analysis for investigating the biological functions of the DEGs, using the clusterProfiler package in R (version 4.1.5). GO terms with $p < 0.05$ were considered to be significantly enriched. The Metascape webservice (<http://metascape.org>) was also used to

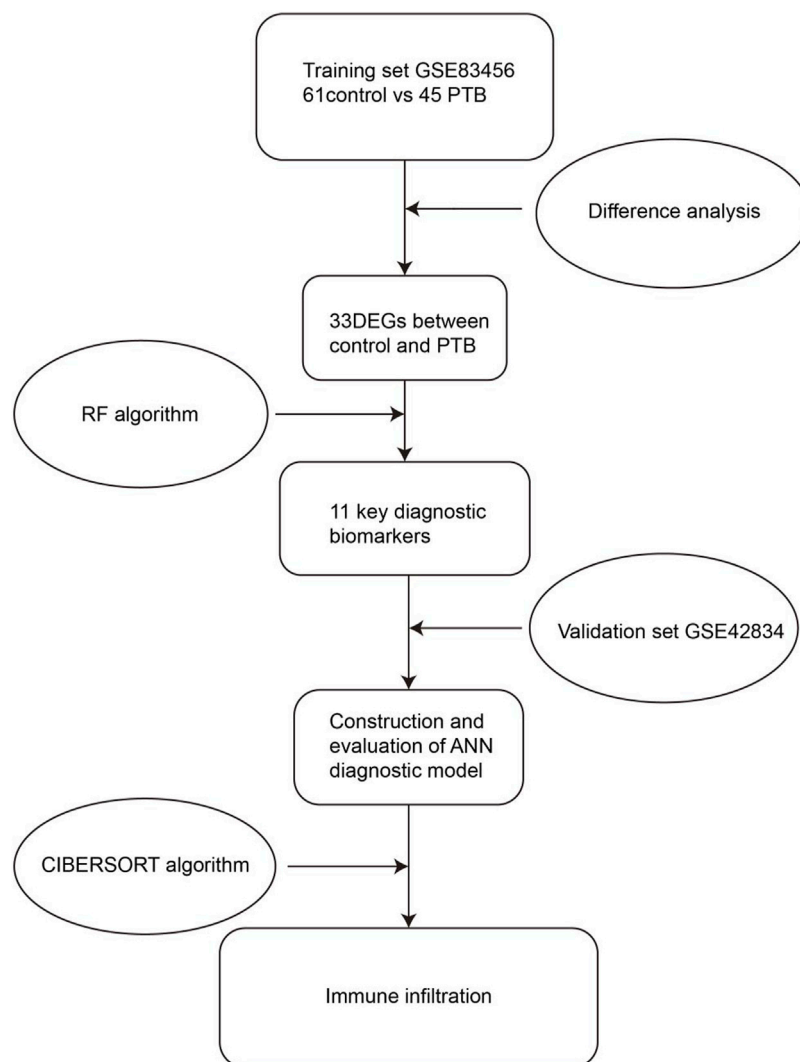


FIGURE 1

Flow chart of the present study. DEGs, differentially expressed genes; RF, random forest; ANN, artificial neural network.

annotate the enriched biological pathways for comprehensive analysis of the biomarkers. The most enriched functions or pathways were subsequently displayed using bubble and bar plots.

Screening significantly enriched DEGs using RF

The DEGs were further screened using the randomForest package in R software. The optimal tree number was first identified based on the best stability and lowest error rate by calculating the error rate of each of the 1–500 trees. We established an RF model based on the optimal tree number for screening the specific PTB genes as candidate biomarkers using the mean decrease in Gini coefficient. In the RF algorithm, a gene importance value greater than 2 is considered to be a common screening criterion, and has been used in other studies on machine learning-based diagnostic models.

Construction and evaluation of an ANN-based diagnostic model

In order to construct an ANN-based diagnostic model, the min-max method was used for normalizing the input data, which were subsequently converted into the “Gene Score” according to the gene expression levels. For instance, the expression of an upregulated gene was denoted as 1 if the expression level was higher than the median expression value across all the samples, or denoted as 0 in other instances. Similarly, the expression of a downregulated gene was generally denoted as 0, or as 1 if the expression level was higher. A neural network-based classification model was subsequently constructed by calculating the weights of the significantly enriched DEGs using the neuralnet package in R (version 4.2.1). A neural network contains an input layer, a hidden layer, and an output layer. In this study, the number of hidden layers was set to 5, and the number of output parameters was set to 2 nodes (contract/segment). Additionally, the AUC value of the training cohort was

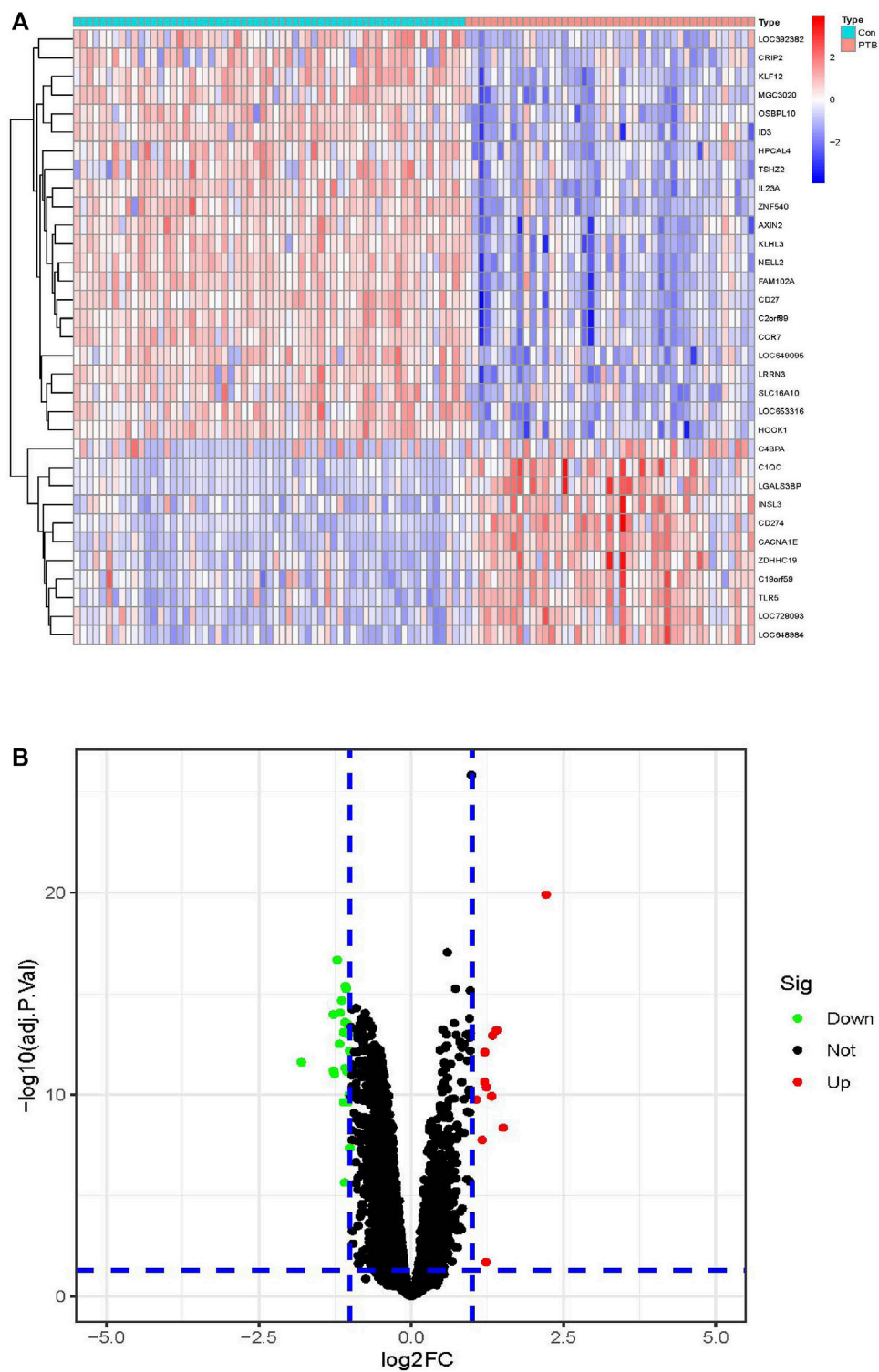


FIGURE 2

Identification of DEGs in the training cohort. **(A)** The heatmap of the 33 DEGs, including 11 upregulated and 22 downregulated ones. PTB were represented by red samples, normal were represented by blue samples. Red blocks indicate high-expressed genes, and blue blocks indicate low-expressed genes. Con, control group; PTB, Pulmonary Tuberculosis. **(B)** Volcano plots of all DEGs in the GSE83456 dataset. Two dotted lines on the X-axis represent the value of $\log_2\text{FC}$ is -1 and 1. The dotted line on the Y-axis represent the adj. p. value is 0.05. Red dots represent high-expressed genes, blue dots represent low-expressed genes and black dots represent not significant changed genes.

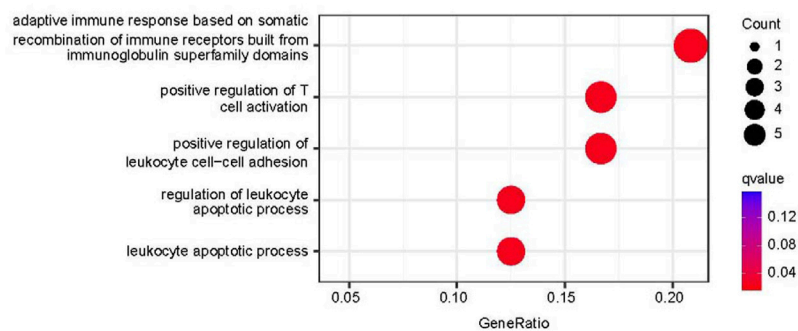


FIGURE 3
Functional enrichment analysis results. Top five enriched GO terms in biological process (BP).

calculated using the pROC package in R (version 4.2.1). The accuracy of the model was also verified using the independent GSE42834 cohort.

Analysis of immune infiltration

CIBERSORT is a deconvolution algorithm that is used for quantifying cell types based on the gene expression profiles, and was used to determine the abundance of 22 types of immune cells in the PTB and control tissues. Using the CIBERSORT algorithm, the immune infiltration landscape in the GSE83456 cohort was comprehensively analyzed, and the differences between the control and PTB groups were depicted using waterfall and correlation plots.

Statistical analyses

The differences in gene expression between the control and PTB samples were compared using Student's t-tests. The categorization effects of the critical biomarkers on the PTB and control specimens were determined using ROC curves and the AUC using the pROC package in R. Statistical analysis was performed using the R software (version 4.2.1) and GraphPad Prism (GraphPad Prism, USA). $p < 0.05$ was considered to be statistically significant, unless otherwise stated.

Results

Data processing and identification of DEGs

The limma package in R was used for identifying the DEGs between the 45 PTB and 61 control samples using the classical Bayesian algorithm, based on the following criteria: $p < 0.05$ and $|\log_2FC| > 1$. A total of 33 DEGs were finally identified, including 11 and 22 DEGs that were significantly upregulated and downregulated, respectively. As depicted in Figure 2A, the expression of these DEGs differed significantly between the PTB and control groups. The results were graphically represented using

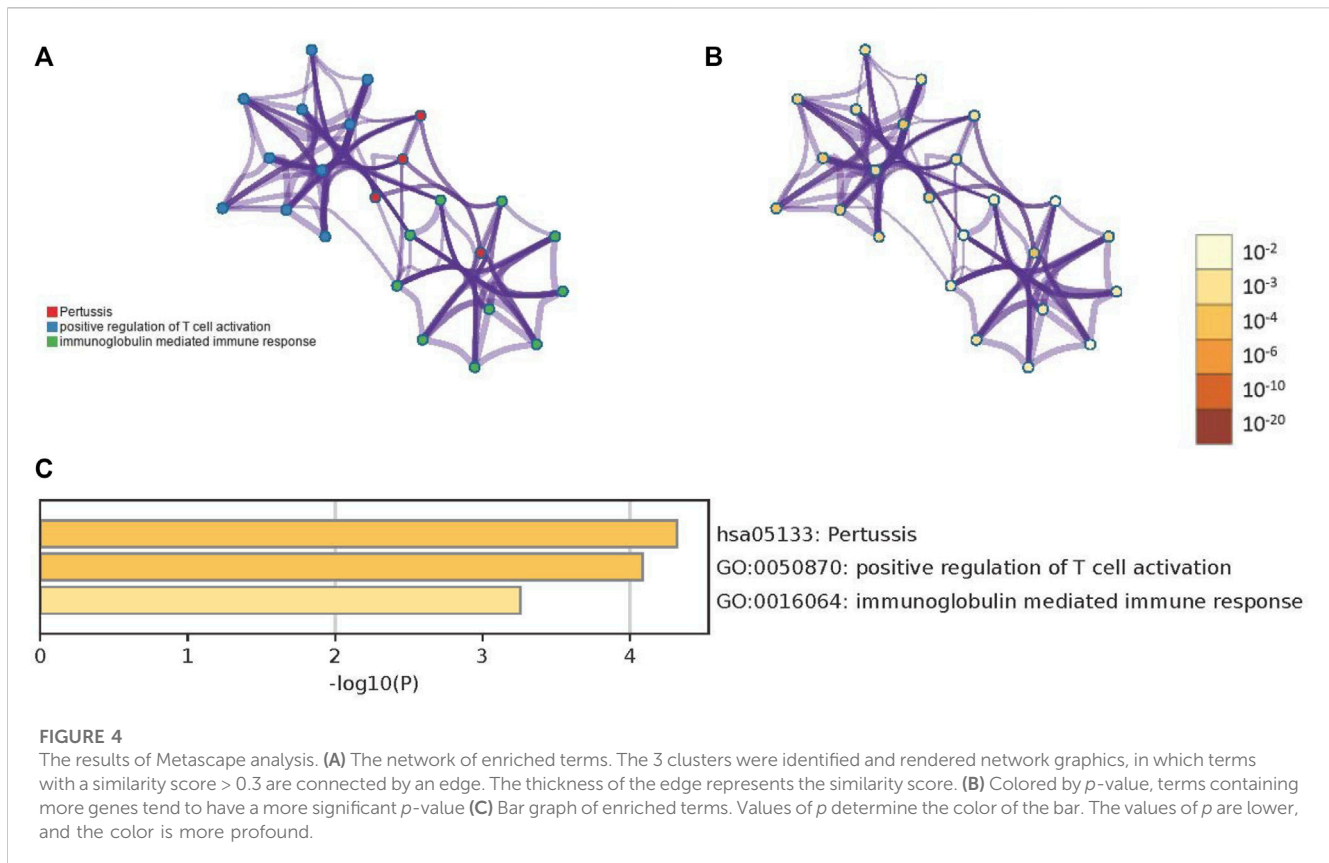
a volcano plot, which further revealed the differences in gene expression and statistical significance of the DEGs (Figure 2B).

Functional enrichment analysis of DEGs

The biological significance of the 33 DEGs in the pathogenesis of PTB was investigated by GO pathway enrichment analysis using the clusterProfiler package in R. The findings revealed that the 33 DEGs were primarily involved in immune-related functions, including adaptive immune response based on somatic recombination of immune receptors comprising immunoglobulin superfamily domains, positive regulation of T cell activation, positive regulation of leukocyte cell-cell adhesion, regulation of leukocyte apoptotic process, and leukocyte apoptotic process. The findings are presented in a bubble plot (Figure 3). The Metascape webserver was also used for annotating the enriched GO terms. The results of Metascape analysis revealed that the three pathways of DEGs were significantly enriched (Figure 4A, B and C).

Screening key genes using an RF classifier

In order to identify the reliable diagnostic biomarkers of PTB, the DEGs were classified using an RF classifier. According to Figure 5A, which depicts the relationship between the RF tree number and the error rate of the model, the trees with the lowest error rate ntrees value (ntrees = 31) were selected. Based on the model accuracy and decreased mean square error, the Gini coefficient method was used for measuring the importance of all the variables. The results of MeanDecreaseGini are provided in Figure 5B. Kruppel-like factor 12 (KLF12) was identified as the most important biomarker. A set of 11 specific biomarkers, including KLF12, interleukin 23 subunit alpha (IL23A), neural EGFL-like 2 (NELL2), Family With Sequence Similarity 102 Member A (FAM102A), Calcium Voltage-Gated Channel Subunit Alpha1 E (CACNA1E), Oxysterol Binding Protein like 10 (OSBPL10), complement component C1q (C1QC), Hook Microtubule Tethering Protein 1 (HOOK1), Chromosome 2 open reading frame 89 (C2orf89), inhibitor of DNA binding 3 (ID3), and Kelch Like Family Member 3 (KLHL3), with



significance >2 were selected as critical biomarkers for further analysis. The heatmap revealed that CACNA1E and C1QC were upregulated in the PTB group, while the remaining 9 genes were downregulated (Figure 5C).

Construction of the ANN model

The weights of each of the biomarkers are provided in Supplementary Table S1. The weights of the 11 biomarkers were analyzed using ANN, based on the gene scores. The ANN model consisted of one input layer, one hidden layer, and one output layer, as depicted in Figure 6A. The input layer included 11 neurons, hidden layer included five neurons and output layer included 2 neurons. The absolute partial derivative of the error function was less than 0.01.

Validation of the ANN model

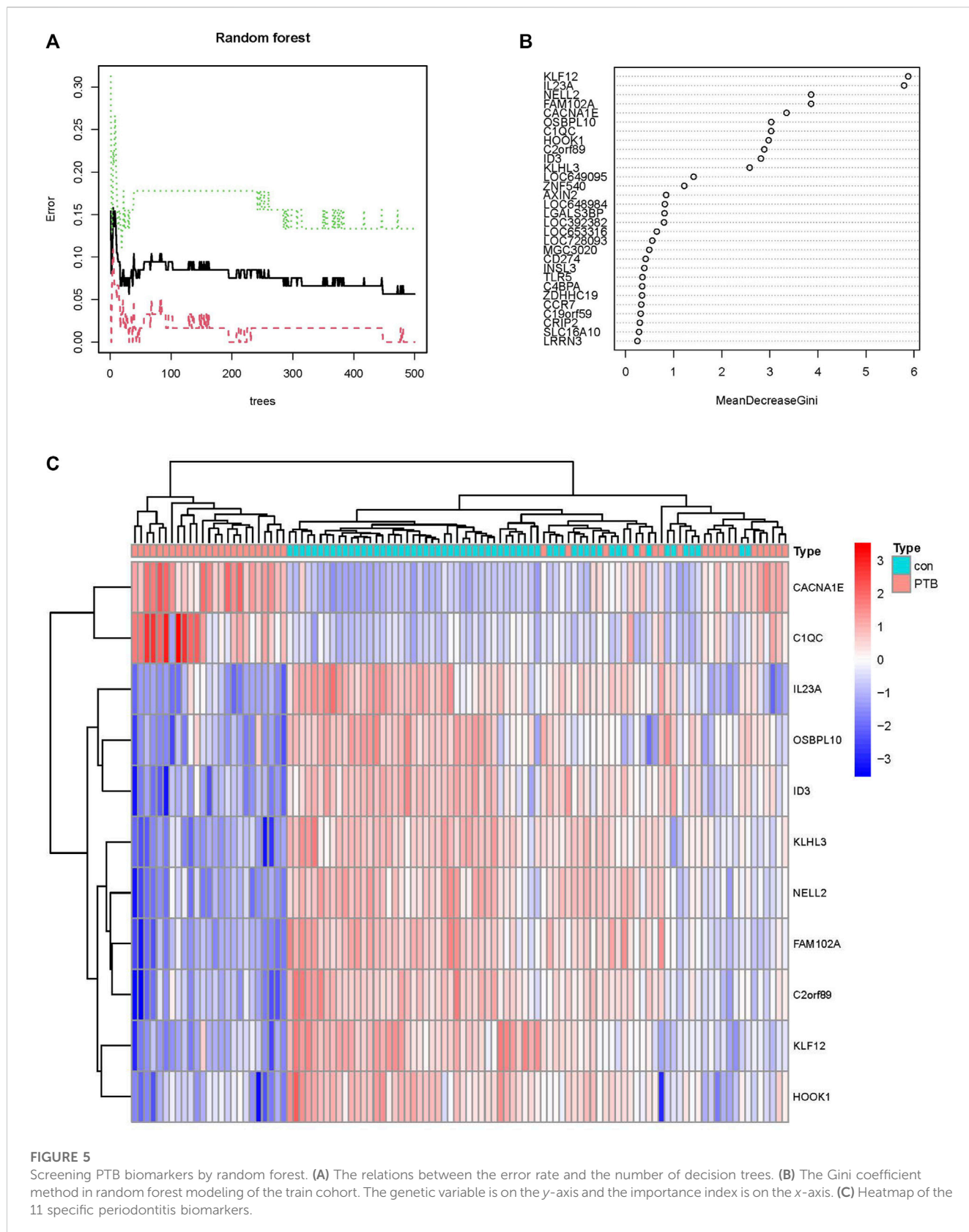
The performance of the ANN model was determined using the pROC package in R, and the AUC of the training GSE83456 cohort was 1.000. This indicated that the ANN model performed exceptionally well in diagnosing PTB (Figure 6B). The ANN model also demonstrated outstanding performance with the independent GSE42834 validation cohort, and the AUC of the validation cohort was determined to be 0.946 (Figure 6C).

Assessment of immune infiltration

The present study further investigated the correlation between the ratios of the 22 types of immunocytes in the PTB and control specimens using the CIBERSORT algorithm. The composition of the immunocytes in the PTB and normal samples and the relationships among the immunocytes are provided in Figure 7A. The findings revealed a positive correlation between the levels of M0 macrophages and monocytes, and between the levels of M0 macrophages and neutrophils. However, there was a negative correlation between the abundance of resting mast cells and activated mast cells, levels of memory B cells and naïve B cells, and the ratio of follicular helper T cells and neutrophils (Figure 7B).

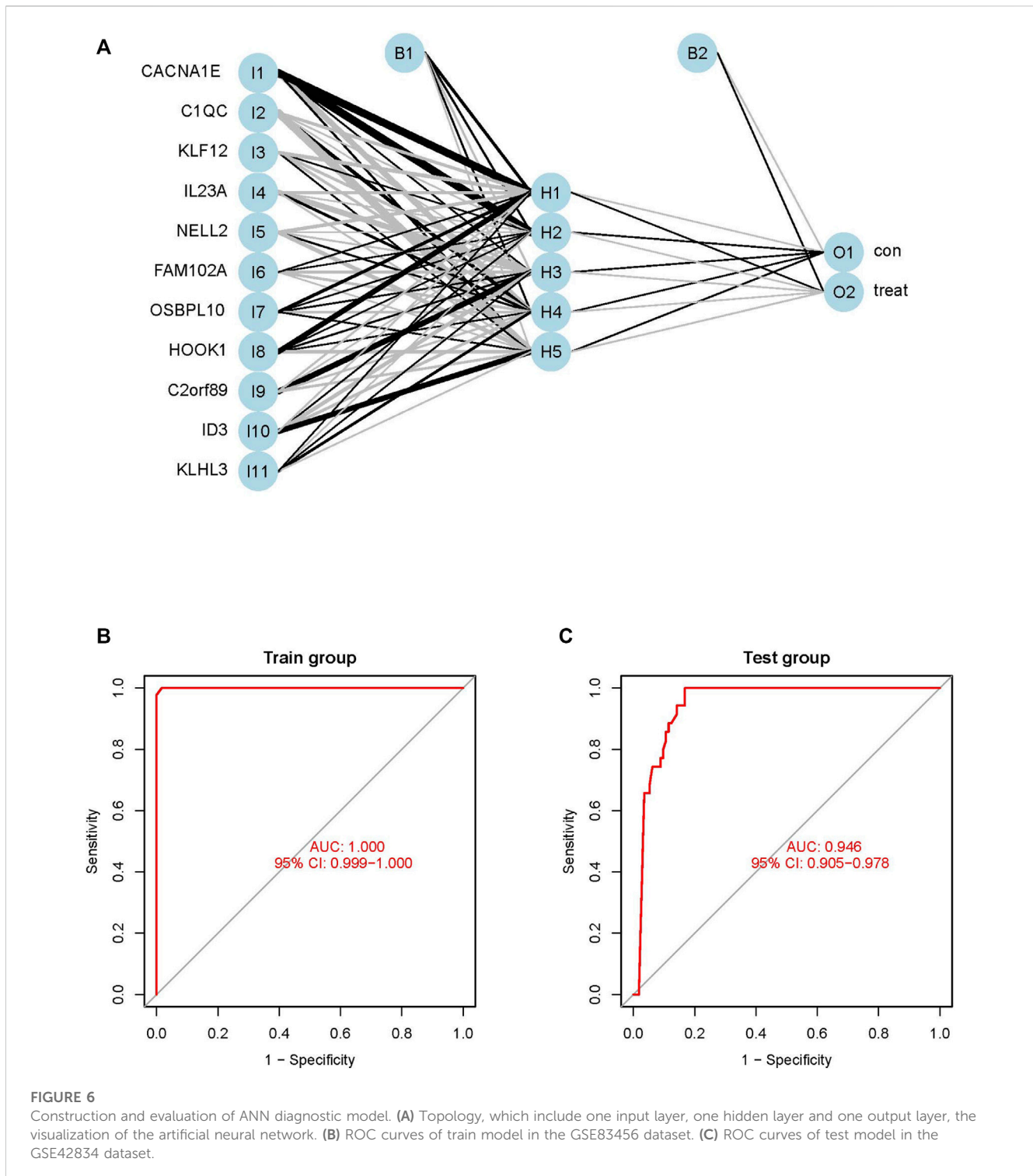
Discussion

The early detection and diagnosis of PTB can reduce its chances of transmission; therefore, identifying specific biomarkers for the prediction of PTB is crucial for controlling disease progression. RF and ANN can be combined for developing reliable diagnostic models for certain diseases, including osteoarthritis and hypertrophic cardiomyopathy (Xie et al., 2020; Li S et al., 2022; Li Z B et al., 2022). RF and ANN are advanced tools for diagnosing PTB, but their main limitation is the necessity for trained and qualified personnel for implementing these tools, as the construction of neural networks, which includes training and testing, is a challenging task. Additionally, the use of statistical tools for diagnosing diseases continues to be a matter of difficulty.



The present study identified 33 DEGs between PTB and control samples in the GSE83456 cohort. A total of 11 candidate genes were identified using an RF classifier, and an ANN algorithm was used for computing the weights of these genes. A classification model was

constructed for the diagnosis of PTB, and a ROC curve was generated for assessing the efficacy of the classification by the ANN model. An independent GSE42834 cohort was used for determining the reliability of the classification model.



The results of enrichment analysis demonstrated that the majority of DEGs were primarily enriched in immune-related functions. It has been reported that T cells are involved in the development of TB, and the activation of T cells enhance resistance to *M. tuberculosis* infections (Feruglio et al., 2017). Leukocytes are also implicated in the inflammatory pathogenesis of TB (Ocana-Guzman et al., 2021). However, adaptive immune responses based on somatic recombination of immune receptors comprising immunoglobulin superfamily domains have not

been previously reported in TB, and may serve as a novel therapeutic target for PTB. Altogether, the findings revealed that these DEG identified herein are positively involved in the immune processes in PTB.

Of the 11 genes screened using the RF classifier, KLF12 (Natarajan et al., 2022), IL23A (Khader et al., 2011), NELL2 (Yang et al., 2015), OSBPL10 (Li et al., 2022), C1QC (Cai et al., 2014), and ID3 (Han et al., 2021) have been identified as candidate biomarkers of TB in previous studies.

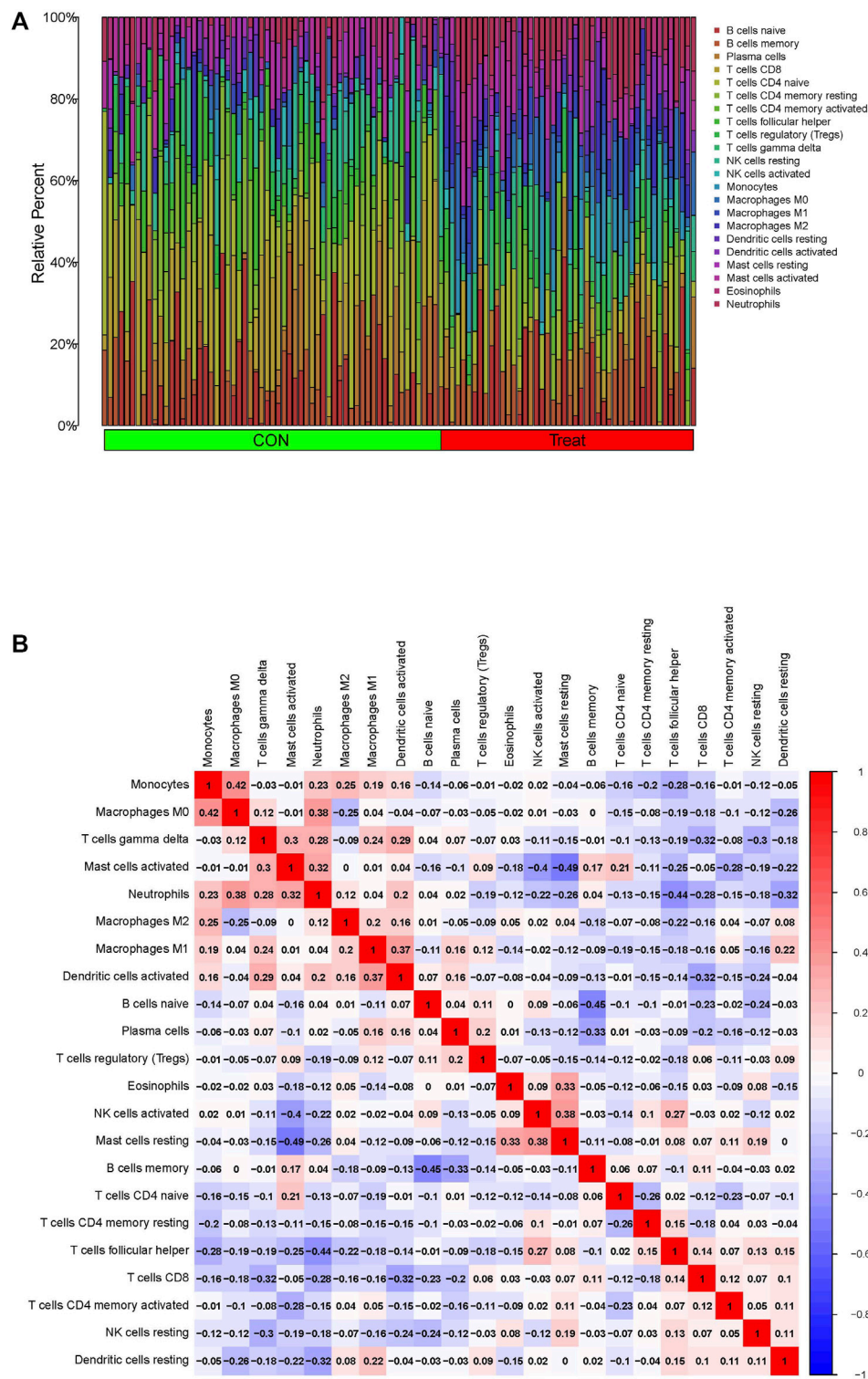


FIGURE 7 Immune infiltration assessment via the CIBERSORT in the GSE83456 dataset. (A) Composition of 22 immunocytes on PTB samples and normal. (B) The relationship among 22 immunocytes are displayed in correlation matrix.

Notably, the present study identified additional five genes that have not been previously shown to be associated with the pathogenesis of PTB. The KLHL3 gene, which is downregulated in PTB, encode proteins that are components of the CullinRING

E3 ubiquitin ligase complex and are involved in the ubiquitin-proteasome system. The complex degrades proteins and also plays an essential role in maintaining cellular functions (Zhang et al., 2022). It has been reported that the ubiquitin-proteasome

system also plays a role in inducing CD8⁺ T cells (Shen et al., 2008). Therefore, the downregulation of KLHL3 may suppress the degradation of proteins that regulate the ubiquitin-proteasome system and subsequently induce CD8⁺ T cells that participate in the pathogenesis of PTB.

The present study revealed that the expression of HOOK1 is downregulated in PTB. A previous study reported that enhancing the interaction between HOOK1 and CD147 may increase the exosomal levels of amyloid- β (Xie et al., 2018). The deposition of amyloid- β has been reported to be associated with tuberculous meningitis (Stroffolini et al., 2021). We therefore speculated that HOOK1 may affect the deposition of amyloid- β to regulate the pathogenesis of PTB. CD147⁺⁺ Tregs cells, a recently described highly suppressive and activated subset of human Tregs, are capable of producing proinflammatory cytokines in TB (Feruglio et al., 2015). These studies collectively suggest that HOOK1 may participate in the pathological processes of PTB *via* multiple pathways.

The CACNA1E protein can mediate the entry of calcium ions into excitable cells and regulate various calcium-dependent processes. Numerous studies have reported that calcium channel blockers have anti-tuberculosis potential (Lee et al., 2015; Song et al., 2015; Lee et al., 2021). Therefore, the upregulation of CACNA1E in PTB may result in the activation of calcium channels and lead to the pathogenesis of PTB.

The present study is the first to identify the association between FAM102A and the pathogenesis of TB. The findings revealed that the expression of FAM102A was downregulated in the samples of PTB in this study. Notably, protein-protein interaction (PPI) analysis with STRING (string-db.org) revealed that the FAM102A protein interacts with NELL2, which has been confirmed as a biomarker of TB. It has been additionally reported that NELL2 plays a crucial role in protecting cells from environments that induce cell death (Kim et al., 2015). The deficiency of NELL2 induces mitochondria-dependent cellular apoptosis and inhibits cellular proliferation by phosphorylating and activating extracellular signal-regulated kinase 1/2 (ERK1/2) (Liu et al., 2021). These findings suggest that FAM102A can function as a biomarker of PTB by interacting with NELL2, and subsequently influence cellular apoptosis and regulate the pathogenesis of PTB.

The C2orf89 protein, also referred to as TRABD2A, could be involved in activating resting CD4⁺ T cells but not activated CD4⁺ T cells. The TRABD2A protein is located on the plasma membrane of resting CD4⁺ T cells and disappears following the activation of T cells (Liang et al., 2019). CD4⁺ T cells produce cytokines, which are vital in controlling *M. tuberculosis* infections (Ferreira et al., 2021). It is therefore likely that the production of cytokines, including interferon (IFN)- γ , by activated CD4⁺ T cells suppresses *M. tuberculosis* infections and downregulates the TRABD2A protein located on the plasma membrane of resting CD4⁺ T cells.

The particularities of our research are combining RF and ANN methods innovatively, and multiple biomarkers combined diagnosis, which showed outstanding results in the predictive power aspect. The AUC of train model and valid model are both greater than 0.9. Compared with several literatures (Manisha Singh et al., 2022; Yu Dong Zhang, 2020) which utilize the chest radiography images to detect Pulmonary Tuberculosis with the help of machine learning tools (CAD, DL, ICNN), our work is analysing biomarkers from peripheral blood biomarkers and constructing diagnostic model for

PTB with the combination of RF and ANN. Although, RF, ANN, or other machine learning had been utilized in diagnosing TB (Dande and Samant, 2018; Orjuela-Canon et al., 2022), combining RF and ANN to diagnose PTB had never been reported. Our samples are both from human blood, we could design the diagnostic kit based on the eleven biomarkers and to detect the blood which sampling from human fingers. It is we choose figure blood sampling rather than sputum smear and X-ray that bring us the diagnostic convenience and safety. However, the present study has certain limitations. Firstly, although our diagnostic model performed well, the number of samples in the training and validation datasets was relatively small. Therefore, independent patient cohorts with a larger sample size are necessary for evaluating the performance of the ANN-based classification model developed herein, and sufficient samples need to be collected from affiliated hospitals for this purpose. Secondly, all the samples were only classified as normal or PTB, which may influence the results of screening; therefore more subtypes of PTB should be considered in future studies. Thirdly, the correlation between the novel biomarkers and the pathogenesis of PTB remain to be determined, and further experimental studies are necessary for elucidating the underlying mechanisms by which the biomarkers regulate the pathogenesis of PTB. Altogether, the model developed herein has high accuracy and excellent diagnostic convenience owing to the use of data obtained from routine blood tests.

Conclusion

Altogether, the present study successfully constructed a novel diagnostic model for PTB. As the diagnostic method is based on peripheral blood tests, a diagnostic kit can be designed based on the 11 biomarkers identified herein, which is highly convenient for the rapid and accurate diagnosis of PTB. The diagnostic model, biomarkers, and the peripheral blood test method discussed herein provide novel insights into the underlying mechanisms and can aid further studies on the clinical diagnosis of PTB. However, further experimental studies are necessary for determining the underlying mechanisms by which the identified biomarkers regulate the pathogenesis of PTB.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This study was supported by Open Program of Health Policy Research Center, Anhui Medical University (2022wszc12).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1094099/full#supplementary-material>

References

- Barac, A., Karimzadeh-Esfahani, H., Pourostadi, M., Rahimi, M. T., Ahmadvan, E., Rashedi, J., et al. (2019). Laboratory cross-contamination of *Mycobacterium tuberculosis*: A systematic review and meta-analysis. *LUNG* 197 (5), 651–661. doi:10.1007/s00408-019-00241-4
- Blankley, S., Graham, C. M., Turner, J., Berry, M. P., Bloom, C. I., Xu, Z., et al. (2016). The transcriptional signature of active tuberculosis reflects symptom status in extra-pulmonary and pulmonary tuberculosis. *PLoS One* 11 (10), e0162220. doi:10.1371/journal.pone.0162220
- Byeon, H. (2019). Developing a random forest classifier for predicting the depression and managing the health of caregivers supporting patients with Alzheimer's disease. *Technol. Health Care* 27 (5), 531–544. doi:10.3233/THC-191738
- Cai, Y., Yang, Q., Tang, Y., Zhang, M., Liu, H., Zhang, G., et al. (2014). Increased complement C1q level marks active disease in human tuberculosis. *PLoS One* 9 (3), e92340. doi:10.1371/journal.pone.0092340
- Curchoe, C. L., Flores-Saiffe, F. A., Mendizabal-Ruiz, G., and Chavez-Badiola, A. (2020). Evaluating predictive models in reproductive medicine. *Fertil. Steril.* 114 (5), 921–926. doi:10.1016/j.fertnstert.2020.09.159
- Dande, P., and Samant, P. (2018). Acquaintance to artificial neural networks and use of artificial intelligence as a diagnostic tool for tuberculosis: A review. *Tuberc. (Edinb)* 108, 1–9. doi:10.1016/j.tube.2017.09.006
- Denkinger, C. M., Schumacher, S. G., Gilpin, C., Korobitsyn, A., Wells, W. A., Pai, M., et al. (2019). Guidance for the evaluation of tuberculosis diagnostics that meet the World health organization (WHO) target product profiles: An introduction to WHO process and study design principles. *J. Infect. Dis.* 220 (220), S91–S98. doi:10.1093/infdis/jiz097
- Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., et al. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *BRIEFINGS Bioinforma.* 14 (6), 671–683. doi:10.1093/bib/bbs046
- Ferreira, C. M., Barbosa, A. M., Barreira-Silva, P., Silvestre, R., Cunha, C., Carvalho, A., et al. (2021). Early IL-10 promotes vasculature-associated CD4+ T cells unable to control *Mycobacterium tuberculosis* infection. *JCI Insight* 6 (21), e150060. doi:10.1172/jci.insight.150060
- Feruglio, S. L., Kvale, D., and Dyrhol-Riise, A. M. (2017). T cell responses and regulation and the impact of *in vitro* IL-10 and TGF-beta modulation during treatment of active tuberculosis. *Scand. J. Immunol.* 85 (2), 138–146. doi:10.1111/sji.12511
- Feruglio, S. L., Tonby, K., Kvale, D., and Dyrhol-Riise, A. M. (2015). Early dynamics of T helper cell cytokines and T regulatory cells in response to treatment of active *Mycobacterium tuberculosis* infection. *Clin. Exp. Immunol.* 179 (3), 454–465. doi:10.1111/cei.12468
- Han, J., Ma, Y., Ma, L., Tan, D., Niu, H., Bai, C., et al. (2021). Id3 and Bcl6 promote the development of long-term immune memory induced by tuberculosis subunit vaccine. *Vaccines (Basel)* 9 (2), 126. doi:10.3390/vaccines9020126
- Jeremiah, C., Petersen, E., Nantanda, R., Mungai, B. N., Migliori, G. B., Amanullah, F., et al. (2022). The WHO Global Tuberculosis 2021 Report - not so good news and turning the tide back to End TB. *Int. J. Infect. Dis.* 124, S26–S29. doi:10.1016/j.ijid.2022.03.011
- Kaforou, M., Wright, V. J., Oni, T., French, N., Anderson, S. T., Bangani, N., et al. (2013). Detection of tuberculosis in HIV-infected and -uninfected african adults using whole blood RNA expression signatures: A case-control study. *PLoS Med.* 10 (10), e1001538. doi:10.1371/journal.pmed.1001538
- Khader, S. A., Guglani, L., Rangel-Moreno, J., Gopal, R., Junecko, B. A., Fountain, J. J., et al. (2011). IL-23 is required for long-term control of *Mycobacterium tuberculosis* and B cell follicle formation in the infected lung. *J. Immunol.* 187 (10), 5402–5407. doi:10.4049/jimmunol.1101377
- Khambati, N., Olbrich, L., Ellner, J., Salgame, P., Song, R., and Bijker, E. M. (2021). Host-based biomarkers in saliva for the diagnosis of pulmonary tuberculosis in children: A mini-review. *Front. Pediatr.* 9, 756043. doi:10.3389/fped.2021.756043
- Khan, M. T., Kaushik, A. C., Ji, L., Malik, S. I., Ali, S., and Wei, D. Q. (2019). Artificial neural networks for prediction of tuberculosis disease. *Front. Microbiol.* 10, 395. doi:10.3389/fmicb.2019.00395
- Khimova, E., Gonzalo, X., Popova, Y., Eliseev, P., Andrey, M., Nikolayevskyy, V., et al. (2022). Urine biomarkers of pulmonary tuberculosis. *Expert Rev. Respir. Med.* 16 (6), 615–621. doi:10.1080/17476348.2022.2090341
- Kim, D. Y., Kim, H. R., Kim, K. K., Park, J. W., and Lee, B. J. (2015). NELL2 function in the protection of cells against endoplasmic reticulum stress. *Mol. CELLS* 38 (2), 145–150. doi:10.14348/molcells.2015.2216
- Kugunavar, S., and Prabhakar, C. J. (2021). Convolutional neural networks for the diagnosis and prognosis of the coronavirus disease pandemic. *Vis. Comput. Ind. Biomed. Art.* 4 (1), 12. doi:10.1186/s42492-021-00078-w
- Lee, C. C., Lee, M. G., Hsu, W. T., Park, J. Y., Porta, L., Liu, M. A., et al. (2021). Use of calcium channel blockers and risk of active tuberculosis disease: A population-based analysis. *HYPERTENSION* 77 (2), 328–337. doi:10.1161/HYPERTENSIONAHA.120.15534
- Lee, M. Y., Lin, K. D., Hsu, W. H., Chang, H. L., Yang, Y. H., Hsiao, P. J., et al. (2015). Statin, calcium channel blocker and Beta blocker therapy may decrease the incidence of tuberculosis infection in elderly Taiwanese patients with type 2 diabetes. *Int. J. Mol. Sci.* 16 (5), 11369–11384. doi:10.3390/ijms160511369
- Li S, S., Feng, Z., Xiao, C., Wu, Y., and Ye, W. (2022). The establishment of hypertrophic cardiomyopathy diagnosis model via artificial neural network and random decision forest method. *Mediat. Inflamm.* 2022, 2024974. doi:10.1155/2022/2024974
- Li Z B, Z. B., Shi, L. Y., Han, Y. S., Chen, J., Zhang, S. Q., Chen, J. X., et al. (2022). Pyridoxal phosphate, pyridoxamine phosphate, and folic acid based on ceRNA regulatory network as potential biomarkers for the diagnosis of pulmonary tuberculosis. *Infect. Genet. Evol.* 99, 105240. doi:10.1016/j.meegid.2022.105240
- Liang, G., Zhao, L., Qiao, Y., Geng, W., Zhang, X., Liu, M., et al. (2019). Membrane metalloprotease TRABD2A restricts HIV-1 progeny production in resting CD4(+) T cells by degrading viral Gag polyprotein. *Nat. Immunol.* 20 (6), 711–723. doi:10.1038/s41590-019-0385-2
- Liu, J., Liu, D., Zhang, X., Li, Y., Fu, X., He, W., et al. (2021). NELL2 modulates cell proliferation and apoptosis via ERK pathway in the development of benign prostatic hyperplasia. *Clin. Sci. (Lond)* 135 (13), 1591–1608. doi:10.1042/CS20210476
- Manisha Singh, G. V. P. S., Handattu Shankaranarayana Akshatha, B. A. A. R., Narasimha, M., Beeraka, A. A. H. G., Akshatha, H. S., Abuhajja, B., et al. (2022). Evolution of machine learning in tuberculosis diagnosis: A review of deep learning-based medical applications. *Electron. J.* 17, 2634. doi:10.3390/electronics1172634
- Morrison, H., and McShane, H. (2021). Local pulmonary immunological biomarkers in tuberculosis. *Front. Immunol.* 12, 640916. doi:10.3389/fimmu.2021.640916
- Natarajan, S., Ranganathan, M., Hanna, L. E., and Tripathy, S. (2022). Transcriptional profiling and deriving a seven-gene signature that discriminates active and latent tuberculosis: An integrative bioinformatics approach. *Genes (Basel)* 13 (4), 616. doi:10.3390/genes13040616
- Nogueira, B., Krishnan, S., Barreto-Duarte, B., Araujo-Pereira, M., Queiroz, A., Ellner, J. J., et al. (2022). Diagnostic biomarkers for active tuberculosis: Progress and challenges. *EMBO Mol. Med.* 14 (12), e14088. doi:10.15252/emmm.202114088
- Ocana-Guzman, R., Tellez-Navarrete, N. A., Ramon-Luing, L. A., Herrera, I., De Ita, M., Carrillo-Alduenda, J. L., et al. (2021). Leukocytes from patients with drug-sensitive and multidrug-resistant tuberculosis exhibit distinctive profiles of chemokine receptor expression and migration capacity. *J. Immunol. Res.* 2021, 6654220. doi:10.1155/2021/6654220
- Orjuela-Canon, A. D., Jutinico, A. L., Awad, C., Vergara, E., and Palencia, A. (2022). Machine learning in the loop for tuberculosis diagnosis support. *Front. Public Health* 10, 876949. doi:10.3389/fpubh.2022.876949

- Shen, J., Hisaeda, H., Chou, B., Yu, Q., Tu, L., and Himeno, K. (2008). Ubiquitin-fusion degradation pathway: A new strategy for inducing CD8 cells specific for mycobacterial HSP65. *Biochem. Biophys. Res. Commun.* 365 (4), 621–627. doi:10.1016/j.bbrc.2007.11.009
- Song, L., Cui, R., Yang, Y., and Wu, X. (2015). Role of calcium channels in cellular antituberculosis effects: Potential of voltage-gated calcium-channel blockers in tuberculosis therapy. *J. Microbiol. Immunol. Infect.* 48 (5), 471–476. doi:10.1016/j.jmii.2014.08.026
- Stroffolini, G., Guastamacchia, G., Audagnotto, S., Atzori, C., Trunfio, M., Nigra, M., et al. (2021). Low cerebrospinal fluid Amyloid- β 1-42 in patients with tuberculous meningitis. *BMC Neurol.* 21 (1), 449. doi:10.1186/s12883-021-02468-2
- Sullivan, J. T., Sulli, C., Nilo, A., Yasmeen, A., Ozorowski, G., Sanders, R. W., et al. (2017). High-throughput protein engineering improves the antigenicity and stability of soluble HIV-1 envelope glycoprotein SOSIP trimers. *J. VIROLOGY* 91 (22), 008622–e917. doi:10.1128/JVI.00862-17
- Sweeney, T. E., Braviak, L., Tato, C. M., and Khatri, P. (2016). Genome-wide expression for diagnosis of pulmonary tuberculosis: A multicohort analysis. *Lancet Respir. Med.* 4 (3), 213–224. doi:10.1016/S2213-2600(16)00048-5
- Wang, N., Chen, J., Xiao, H., Wu, L., Jiang, H., and Zhou, Y. (2019). Application of artificial neural network model in diagnosis of Alzheimer's disease. *BMC Neurol.* 19 (1), 154. doi:10.1186/s12883-019-1377-4
- Xie, J. C., Ma, X. Y., Liu, X. H., Yu, J., Zhao, Y. C., Tan, Y., et al. (2018). Hypoxia increases amyloid-beta level in exosomes by enhancing the interaction between CD147 and Hook1. *Am. J. Transl. Res.* 10 (1), 150–163.
- Xie, N. N., Wang, F. F., Zhou, J., Liu, C., and Qu, F. (2020). Establishment and analysis of a combined diagnostic model of polycystic ovary syndrome with random forest and artificial neural network. *Biomed Res. Int.* 2020, 2613091. doi:10.1155/2020/2613091
- Yang, Y., Mu, J., Chen, G., Zhan, Y., Zhong, J., Wei, Y., et al. (2015). iTRAQ-based quantitative proteomic analysis of cerebrospinal fluid reveals NELL2 as a potential diagnostic biomarker of tuberculous meningitis. *Int. J. Mol. Med.* 35 (5), 1323–1332. doi:10.3892/ijmm.2015.2131
- Yu Dong Zhang, D. R. N. X., Nayak, D. R., Zhang, X., and Wang, S. H. (2020). Diagnosis of secondary pulmonary tuberculosis by an eight-layer improved convolutional neural network with stochastic pooling and hyperparameter optimization. *J. Ambient Intell. Humaniz. Comput.*, 1, doi:10.1007/s12652-020-02612-9
- Zak, D. E., Penn-Nicholson, A., Scriba, T. J., Thompson, E., Suliman, S., Amon, L. M., et al. (2016). A blood RNA signature for tuberculosis disease risk: A prospective cohort study. *LANCET* 387 (10035), 2312–2322. doi:10.1016/S0140-6736(15)01316-1
- Zhang, W., Zhang, Q., Che, L., Xie, Z., Cai, X., Gong, L., et al. (2022). Using biological information to analyze potential miRNA-mRNA regulatory networks in the plasma of patients with non-small cell lung cancer. *BMC CANCER* 22 (1), 299. doi:10.1186/s12885-022-09281-1
- Zhou, Y., Yu, Z., Liu, L., Wei, L., Zhao, L., Huang, L., et al. (2022). Construction and evaluation of an integrated predictive model for chronic kidney disease based on the random forest and artificial neural network approaches. *Biochem. Biophys. Res. Commun.* 603, 21–28. doi:10.1016/j.bbrc.2022.02.099