Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable machine learning-assisted differentiation of brucellosis from tuberculosis via routine blood biomarkers

Wen Song
Tumor Hospital Affiliated to Xinjiang Medical UniversityState Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Laboratory Center

Jinping Zhou
Tumor Hospital Affiliated to Xinjiang Medical UniversityState Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Laboratory Center

Ruixue Xu
Tumor Hospital Affiliated to Xinjiang Medical UniversityState Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Laboratory Center

Junjie Deng
Xinjiang Medical University Affiliated Traditional Chinese Medicine Hospital

Junjie Ma
Xinjiang Medical University Affiliated Traditional Chinese Medicine Hospital

Zulihumaer Yasheng
Xinjiang Medical University Affiliated Traditional Chinese Medicine Hospital

Xin Xiang
Xinjiang Medical University Affiliated Traditional Chinese Medicine Hospital

Lidan Guo
Xinjiang Medical University Affiliated Traditional Chinese Medicine Hospital

Yumei Liu
Xinjiang Medical University Affiliated Traditional Chinese Medicine Hospital

Xiumin Ma
maxiumin1210@sohu.com

Tumor Hospital Affiliated to Xinjiang Medical UniversityState Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Laboratory Center

---

Research Article

Keywords: Brucellosis, Tuberculosis, Machine learning, Biomarkers, Clinical Decision Support Tool

Posted Date: September 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10141297/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable machine learning-assisted differentiation of brucellosis from tuberculosis via routine blood biomarkers

Wen Song[1, 2†], Jinping Zhou[2†], Ruixue Xu[1,2†], Junjie Deng[2], Junjie Ma[2],Zulihumaer Yasheng[2],Xin Xiang[2], Lidan Guo[2],Yumei Liu[2]*, Xiumin Ma[1]*

1. State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Laboratory Center, Tumor Hospital Affiliated to Xinjiang Medical University, Urumqi 830000, P. R. China

2. Clinical Laboratory Center, Xinjiang Medical University Affiliated Traditional Chinese Medicine Hospital, Urumqi 830000, P. R. China

† These authors contributed equally to this work and should be considered co-first authors.

* Corresponding Author: Yumei Liu (49644958@qq.com); Xiumin Ma (maxiumin1210@163.com).

Abstract

Background: To enable low-cost differential diagnosis in primary care settings, we developed a machine-learning model integrating routine clinical indicators and occupational exposure history to discriminate between brucellosis and tuberculosis. Key predictive features were identified via feature selection, and a parsimonious model was subsequently constructed. The model's generalizability was rigorously validated

through internal cross-validation and external testing to provide an affordable, practical auxiliary tool for resource-limited primary healthcare.

Methods: A retrospective cohort of 273 patients from the Traditional Chinese Medicine Hospital affiliated with Xinjiang Medical University was enrolled. After 1:1 propensity score matching, 180 patients (90 with brucellosis, 90 with tuberculosis) were included. Thirty-five clinical variables and occupation were recorded. Four models—logistic regression, support vector machine, random forest (RF), and XGBoost—were developed using 10-fold stratified cross-validation. SHAP analysis identified core features. A three-variable simplified RF model was constructed and validated on an independent temporal test set (n = 42).

Results: Among the evaluated algorithms, random forest (RF) achieved the best internal cross-validation performance, with an AUC of 0.9056 ± 0.0697, whereas XGBoost exhibited greater variability. SHAP analysis identified occupation, alanine aminotransferase (ALT), and γ-glutamyl transferase (GGT) as the most influential predictors. A simplified RF model incorporating only these three variables yielded a cross-validated AUC of 0.834 ± 0.1052. Compared with the full-feature model, the simplified version retained 92.09% of the original AUC and 97.26% of the sensitivity. Although a minor decrease in accuracy was observed, the McNemar test revealed no statistically significant difference between the two models (P = 0.0776) at the α = 0.05 level. Decision curve analysis showed comparable clinical net benefit. In the external test set, the simplified model attained an AUC of 0.8265 (95% CI: 0.696 – 0.957), with a sensitivity of 85.71% and a specificity of 71.43%.

Conclusion: The three-variable simplified random forest model — incorporating occupation, GGT, and ALT — exhibits robust discrimination in distinguishing brucellosis from tuberculosis. It demonstrates reliable generalizability in external validation and represents a simple, cost-effective diagnostic aid for primary care settings.

Keywords: Brucellosis; Tuberculosis;Machine learning; Biomarkers; Clinical Decision Support Tool

Background

Brucellosis is a zoonotic infection caused by the bacterium Brucella species Along with tuberculosis, it remains highly endemic in China's pastoral regions, including Xinjiang[1, 2]. Clinically, the two diseases share considerable overlap, typically manifesting with nonspecific systemic symptoms—such as intermittent fever, night sweats, fatigue, arthralgia, and weight loss—and both can involve multiple organ systems, notably the liver and the osteoarticular system, rendering early and accurate differentiation challenging[2–4]. Routine laboratory inflammatory markers in affected patients often show nonspecific derangements, further confounding the differential diagnosis. These diagnostic difficulties are particularly pronounced in primary healthcare settings, where the initial misdiagnosis rate exceeds 30%, and the average time to a definitive diagnosis may surpass 40 days[5]. The management and control of infectious diseases rely heavily on timely and accurate diagnosis. Bacterial culture

techniques and serological methods are the primary diagnostic tools. These methods also contribute to monitoring programs, population screening, and planning, control, and eradication strategies worldwide[6].

Current conventional methods for the differential diagnosis of brucellosis and pulmonary tuberculosis have significant limitations and fail to meet the need for rapid, accurate first-line diagnosis in primary care[6, 7]. Although pathogen culture is the "gold standard" for the clinical diagnosis of both diseases, its long turnaround time and low positivity rate severely restrict early diagnostic performance. Studies show that the blood culture positivity rate for brucellosis is only 10%—30%, and the sputum culture positivity rate for pulmonary tuberculosis is less than 50%, reflecting low overall detection efficiency[8]. Furthermore, mycobacterial culture requires demanding laboratory conditions, equipment, and technical expertise, making it unavailable in most primary hospitals and hindering widespread access to pathogen-based diagnosis[5, 9, 10].Serological testing also has shortcomings: cross-reactions between antigens of different bacterial species (e.g.,, *Brucella* and *Yersinia enterocolitica O:9*) can lead to false-positive results, directly affecting diagnostic accuracy and increasing the risk of misdiagnosis and missed diagnosis. Molecular assays, while offering high sensitivity and partially compensating for the deficiencies of traditional culture, are expensive and require high-grade laboratory facilities, sophisticated instruments, and skilled personnel, making them difficult to implement in primary settings[5, 11]; moreover, no commercial kit is currently available for the in vitro diagnosis of human brucellosis. Advanced imaging techniques such as CT and MRI are not widely accessible due to

equipment limitations and high costs, and they are difficult to deploy in resource-constrained regions. Although imaging may provide some discriminatory value for patients with osteoarticular involvement in brucellosis, its overall clinical utility remains limited[12]. Taken together, these diagnostic obstacles—including limited imaging accessibility, delays in microbiological culture, and biosafety concerns—collectively contribute to poor patient outcomes and increased healthcare burdens.

Routine hematological and liver function tests, along with other standard laboratory parameters, are part of the regular workup for patients with suspected *Brucella* or *tuberculosis* infection[6, 7], and are available in almost all healthcare settings. These blood biomarkers are particularly well-suited for building machine learning-based diagnostic models for several reasons: first, they are universally accessible and provide rapid results, supporting fast clinical decisions; second, they are more cost-effective than advanced imaging or specialised assays; and third, they can capture the host systemic immune response patterns that are essential for distinguishing brucellosis from tuberculosis [6, 13], thereby addressing diagnostic challenges in infectious diseases.

Compared with the many limitations of conventional diagnostic methods, machine learning (ML) algorithms can effectively uncover complex nonlinear associations and feature interactions in clinical data, transcending the bottlenecks of manual data analysis. ML offers a new technical pathway for constructing efficient and accurate auxiliary differential diagnostic models, revolutionizing the diagnosis of infectious diseases [13 – 15]. Previous studies have applied ML to the differential diagnosis of

spinal infections, achieving accurate discrimination between tuberculous spondylitis and brucellosis spondylitis using imaging data, with model AUC values ranging from 0.85 to 0.92, thereby demonstrating the potential value of ML in differentiating these two diseases [7]. However, current ML research specifically targeting the differential diagnosis of brucellosis and tuberculosis still has notable shortcomings that limit its applicability to primary care. Specifically, most studies have used a single ML algorithm, lacking systematic comparisons and optimal selection across multiple algorithms. Moreover, many studies rely heavily on radiomics features to build models that are not aligned with the diagnostic capabilities of primary healthcare facilities, thereby hindering clinical deployment and scaling. In addition, existing models often suffer from poor interpretability ("black box" problems), failing to clearly elucidate the rationale for diagnostic decisions, which limits clinician acceptance and severely impedes clinical translation and practical application.

There is an urgent need for better differential diagnosis of brucellosis and tuberculosis. The delays and biosafety concerns associated with conventional methods further heighten this urgency. This study aims to systematically compare the performance of four mainstream machine learning algorithms for diagnosing brucellosis and tuberculosis using routine clinical biomarkers. It will quantify the independent contributions of clinical features using SHAP interpretability analysis. We will further identify core discriminatory features and build a streamlined prediction model. Finally, we will validate the model's generalizability using a temporally

independent external test set. Our goal is to establish a low-cost, easy-to-implement intelligent diagnostic tool suitable for use in primary healthcare settings.

Materials and methods

Study population and data collection

This study was a single-center, retrospective, observational investigation. Electronic medical records of patients diagnosed with brucellosis or tuberculosis at the Affiliated Traditional Chinese Medicine Hospital of Xinjiang Medical University between January 2021 and February 2026 were reviewed. All data, including demographic information, chief complaints, clinical signs and symptoms, past medical history, comorbidities, and admission laboratory results, were extracted from the hospital's unified electronic medical record system. The study protocol was approved by the Ethics Committee of Xinjiang Uygur Autonomous Region Traditional Chinese Medicine Hospital (approval No. 2026XE0115). Owing to the retrospective design and the complete anonymization of patient data, a waiver of written informed consent was granted.

Inclusion criteria for the brucellosis group were as follows: patients were required to meet the confirmed diagnostic criteria outlined in the Chinese National Health Industry Standard WS 269-(2019) Diagnosis of brucellosis and fulfill at least one of the following: (a) positive culture of *Brucella* species from blood or bone marrow; (b) standard tube agglutination test (SAT) titer ≥ 1:100 accompanied by compatible clinical symptoms and/or epidemiological exposure history; or (c) positive Rose Bengal plate

test together with an SAT titer ≥ 1:100 in the presence of corresponding clinical symptoms and/or epidemiological exposure history. Additionally, patients had to have complete or near-complete clinical and laboratory data and be 18 years of age or older. Joint/bone involvement was defined based on documented arthralgia, spinal pain, arthritis, or spondylitis in the medical records. Smoking history was defined as current smoking for at least one year, and alcohol consumption history as drinking at least once per week for one year or more. Comorbidities included hypertension and diabetes, both determined according to explicit clinical diagnoses recorded in the medical charts.

Inclusion criteria for the tuberculosis group were as follows: patients were required to meet the diagnostic criteria of the Chinese National Health Industry Standard WS 288-(2017) Diagnosis of Pulmonary Tuberculosis (extrapulmonary tuberculosis was diagnosed according to the Chinese clinical guidelines for tuberculosis diagnosis) and fulfill at least one of the following: (a) positive acid-fast bacilli smear and/or positive culture for the *Mycobacterium tuberculosis* complex from sputum, bronchoalveolar lavage fluid, or pleural effusion; (b) positive nucleic acid amplification test for M. tuberculosis; or (c) histopathological findings of a typical tuberculous granuloma with caseous necrosis. Additionally, patients were required to have complete clinical and laboratory data comparable to those of the brucellosis group and to be 18 years of age or older.

The common exclusion criteria for both groups were as follows: patients were excluded if they presented with any of the following: a missing rate of core clinical variables exceeding 30%; other active infectious diseases, including but not limited to

viral hepatitis or HIV infection; a confirmed malignant tumor; pregnancy or breastfeeding; or a confirmed autoimmune disease, including but not limited to systemic lupus erythematosus or ankylosing spondylitis.

Candidate Features

A total of 35 candidate features were collected. Demographic characteristics included occupation, sex, and age. Clinical symptoms and lifestyle features included smoking history, alcohol consumption history, and comorbidities. Laboratory features comprised 29 continuous variables, all derived from fasting venous blood samples collected within 24 hours of admission and prior to any antimicrobial therapy. These included nine liver function indicators: ALT, AST, AST/ALT ratio, alkaline phosphatase (ALP), GGT, total protein (TP), albumin (ALB), globulin (GLB), and albumin/globulin ratio (A/G); four kidney function indicators: blood urea nitrogen (BUN), creatinine (CREA), uric acid (URIC), and BUN/creatinine ratio (BUN/CREA); nine complete blood count indicators: white blood cell count (WBC), absolute neutrophil count (N), absolute lymphocyte count (L), absolute monocyte count (M), absolute eosinophil count (E), absolute basophil count (BASO), red blood cell count (RBC), hemoglobin (HB), and platelet count (PLT); four inflammatory and metabolic indicators: high-sensitivity C-reactive protein (hsCRP), procalcitonin (PCT), and blood glucose (GLU); and three cardiac enzyme indicators: creatine kinase (CK), creatine kinase-MB isoenzyme (CK-MB), and lactate dehydrogenase (LDH). The hospital

clinical laboratory performed all tests in accordance with standardized operating procedures and conducted routine quality control.

Propensity Score Matching (PSM)

To minimize confounding bias due to age and sex, 1:1 nearest-neighbor propensity score matching without replacement was performed. A logistic regression model was constructed with group assignment as the dependent variable and age and sex as independent variables to estimate propensity scores. The caliper width was set to 0.5 standard deviations of the logit of the propensity score. Standardized mean difference (SMD) was used to assess balance after matching, with an SMD <0.1 indicating good balance. After matching, 90 matched pairs (n = 180) were retained for subsequent analyses.

Machine Learning Model Development and Validation

A stratified ten-fold cross-validation was implemented for model training and evaluation. The mean area under the curve (AUC) served as the optimization objective. Hyperparameters were determined via a grid search; following tuning, the final model was retrained on the complete dataset (n = 180). Model performance was assessed using AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), F1 score, and Kappa. A random seed of 42 was fixed. For logistic regression, variable selection was initially conducted using LASSO with ten-fold cross-validation, followed by Firth's penalized logistic regression to mitigate separation issues in small samples. For the

support vector machine (SVM), recursive feature elimination (RFE) with a random forest base learner was applied, guided by accuracy via 10-fold cross-validation. The number of features was capped at 10 or fewer. A linear kernel was employed, and Z-score normalization was performed independently per fold. The cost parameter C was optimized via a grid search over the range 2^{-5} to 2^{5}. For random forest (RF), all 35 features were utilized. The number of trees was fixed at 300, with a minimum node size of 10. The mtry parameter was optimized via grid search over {2, 4, 6, 8, 10} based on mean AUC. For XGBoost, all features were used, and the model was evaluated using stratified ten-fold cross-validation. SHAP analysis was performed on the more robust models. The mean absolute SHAP values were used for feature ranking, and summary plots (i.e., bee swarm plots) and dependence plots were generated. Key distinguishing features were selected based on SHAP importance to construct a parsimonious model, following the same modeling strategy as the full-feature model. Performance retention rate was defined as (simplified model metric / full-feature model metric) × 100%. The cross-fold AUC values of the two models were compared, and decision curve analysis was conducted.

Temporal external validation

A temporal external validation dataset was assembled from the same institution, covering the period from January 2018 to December 2020, with no overlap with the primary dataset. The same inclusion and exclusion criteria were applied, and data collection was limited to the three variables required for the simplified model (GGT,

ALT, and occupation). A total of 42 patients (21 with brucellosis and 21 with tuberculosis) were enrolled. The external validation set was not used for model training, hyperparameter optimization, or feature selection. Using the simplified random forest model trained on the full primary dataset and a default classification threshold of 0.5, the probability of brucellosis was predicted for each patient in the external validation set. Model performance was assessed using the area under the curve (AUC; with 95% confidence intervals estimated via 1,000 bootstrap resampling iterations), as well as accuracy, sensitivity, specificity, positive predictive value (PPV), F1 score, and Kappa coefficient.

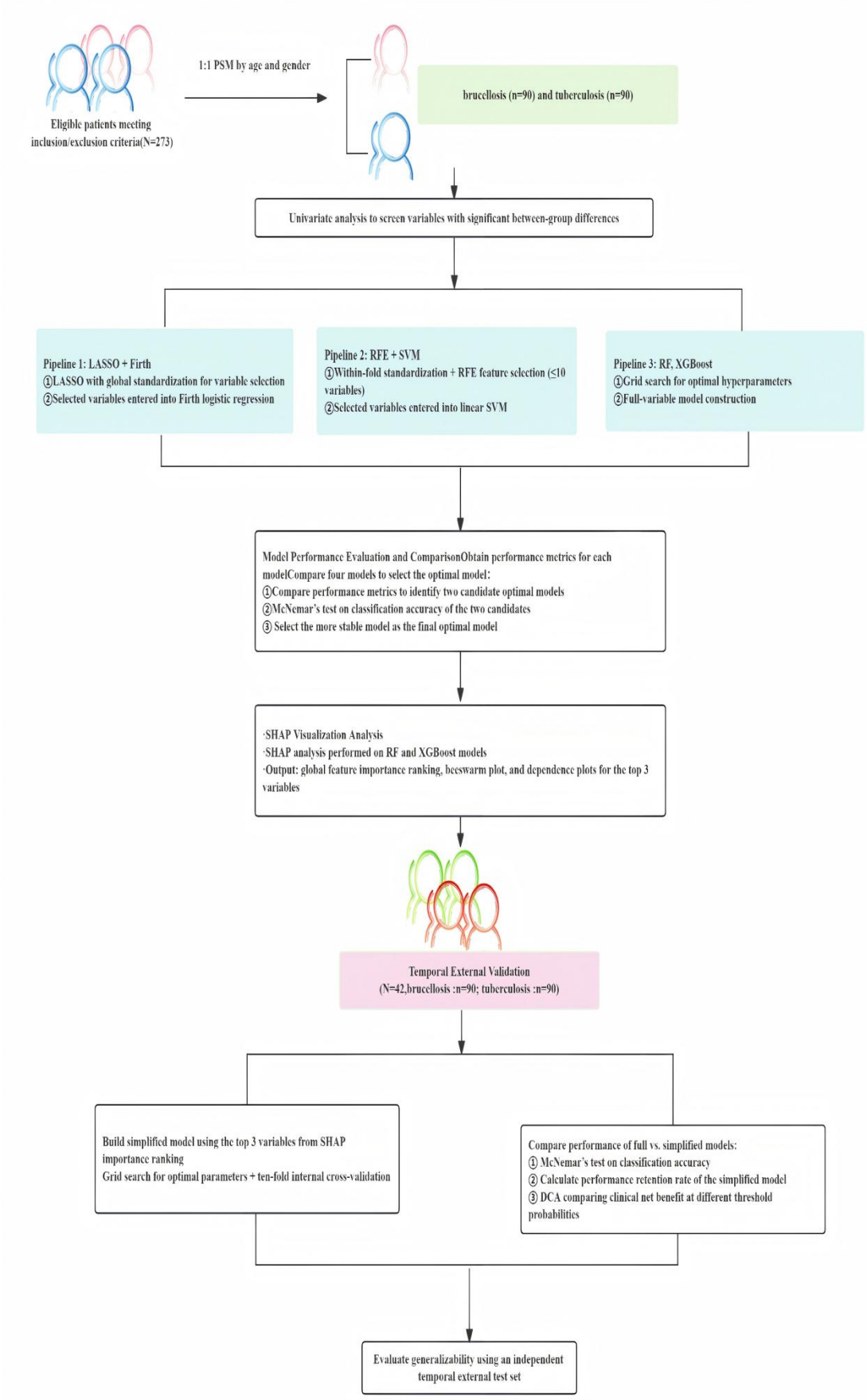
Statistical analysis

All statistical analyses, machine learning modeling, and visualizations were performed using R version 4.2.3. The main R packages employed included glmnet (for LASSO regression), brglm2 (for Firth's penalized logistic regression), e1071 (for support vector machines), randomForest (for random forests), XGBoost (for extreme gradient boosting), caret (for cross-validation and hyperparameter tuning), pROC (for ROC analysis), ggplot2 (for visualization), rmda (for decision curve analysis), and SHAP . Continuous variables that were not normally distributed were presented as medians with interquartile ranges; categorical variables were presented as frequencies (percentages). Between-group comparisons were conducted using the Wilcoxon rank-sum test for continuous variables and the chi-square test or Fisher's exact test for categorical variables. A two-sided P value <0.05 was considered statistically significant.

Results

Baseline characteristics of the study population

As shown in the study flow diagram (Figure 1), 273 patients were initially enrolled in this retrospective analysis. This group included 105 with brucellosis and 168 with tuberculosis. After 1:1 propensity score matching (PSM), 180 patients were included for subsequent analyses. Before matching, the tuberculosis group was older (62.31 ± 15.23 years) and had more females (61.3%, 103/168). The brucellosis group was younger (55.34 ± 10.96 years) and predominantly male (75.2%, 79/105). The groups differed significantly in age (P<0.001) and sex distribution (P<0.001), showing baseline imbalances. After 1:1 PSM, the groups became balanced and comparable for sex (P = 0.7462) and age (P = 0.7387) (Supplementary Table 1).


Eligible patients meeting inclusion/exclusion criteria(N=273)
1:1 PSM by age and gender
brucellosis (n=90) and tuberculosis (n=90)
Univariate analysis to screen variables with significant between-group differences
Pipeline 1: LASSO + Firth
①LASSO with global standardization for variable selection
②Selected variables entered into Firth logistic regression
Pipeline 2: RFE + SVM
①Within-fold standardization + RFE feature selection (≤10 variables)
②Selected variables entered into linear SVM
Pipeline 3: RF, XGBoost
①Grid search for optimal hyperparameters
②Full-variable model construction
Model Performance Evaluation and ComparisonObtain performance metrics for each modelCompare four models to select the optimal model:
①Compare performance metrics to identify two candidate optimal models
②McNemar's test on classification accuracy of the two candidates
③ Select the more stable model as the final optimal model
·SHAP Visualization Analysis
·SHAP analysis performed on RF and XGBoost models
·Output: global feature importance ranking, beeswarm plot, and dependence plots for the top 3 variables
Temporal External Validation
(N=42,brucellosis :n=90; tuberculosis :n=90)
Build simplified model using the top 3 variables from SHAP importance ranking
Grid search for optimal parameters + ten-fold internal cross-validation
Compare performance of full vs. simplified models:
① McNemar's test on classification accuracy
② Calculate performance retention rate of the simplified model
③ DCA comparing clinical net benefit at different threshold probabilities
Evaluate generalizability using an independent temporal external test set

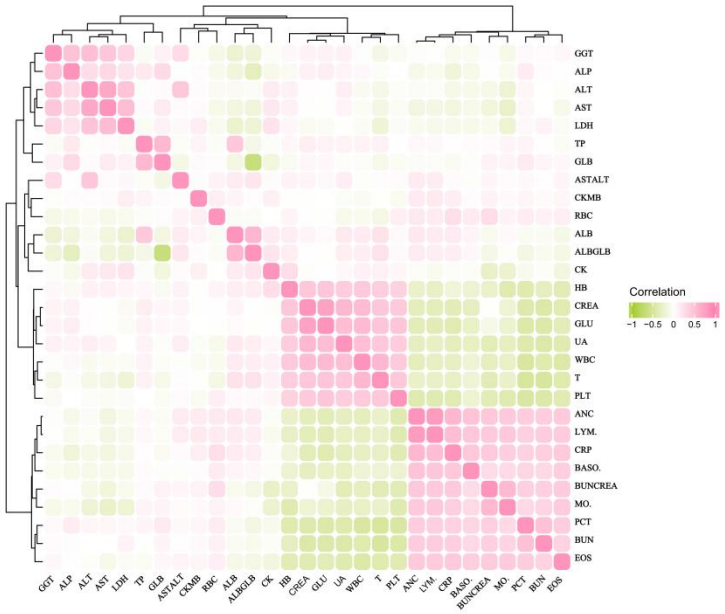
275

No significant intergroup differences were observed between matched patients regarding age, sex, alcohol consumption, smoking status, hypertension, diabetes mellitus, and most laboratory parameters (WBC, ANC, LYM, MO, EOS, PLT, Hb, RBC, PCT, hsCRP) ($P > 0.05$) (Table 1).Some key clinical features did differ. The tuberculosis group had higher body temperature ($P < 0.001$). The brucellosis group had more joint and vertebral lesions (62.2% vs. 32.2%, $P < 0.001$). Occupational distribution also differed ($P < 0.001$): more farmers/herders in the brucellosis group (53.3%) and more retirees in the tuberculosis group (50.0%). In terms of blood chemistry, the brucellosis group had higher uric acid ($P = 0.024$), GGT, ALP, ALT, AST, and LDH ($P < 0.001$), but lower albumin, globulin, and albumin/globulin ratio ($P = 0.001$, 0.002, and <0.001). In summary, brucellosis involved more osteoarticular and hepatic involvement, while tuberculosis showed more fever and higher albumin. These differences support further differential diagnostic analysis.

Correlation analysis of clinical indicators

To assess the degree of multicollinearity among the 29 continuous clinical variables and to prevent highly correlated variables from interfering with model stability and feature importance estimation, we calculated Pearson correlation coefficients for all continuous variables and constructed a correlation clustering heatmap (Figure 2). The heatmap revealed that the absolute correlation coefficients of all continuous variables were below 0.8, indicating no severe multicollinearity. The variables displayed clear functional clustering: liver function indices (ALT, AST, GGT, ALP) were positively

correlated with one another, with the highest correlation observed between ALT and AST (r = 0.72), which is consistent with the pathophysiological pattern of their concomitant elevation during hepatocellular injury. Routine haematological inflammatory markers and renal function indices also exhibited moderate positive correlations within their respective clusters.



Construction and performance evaluation of machine learning models

Hyperparameters are critical determinants of model fitting capacity and generalization performance. In this study, grid search combined with ten-fold cross-validation was employed to optimize hyperparameters for all models, thereby preventing data leakage and ensuring the generalizability of the optimal parameters (Supplementary Table 2).

A multidimensional evaluation framework was adopted to comprehensively assess the performance of the four candidate models, ensuring scientific rigor and clinical applicability in model selection. Univariate analysis (Table 1) was used solely to

evaluate between-group differences for each variable, reflecting the independent discriminatory capacity of individual variables but unable to capture potential interactions or synergistic effects among variables. Relying solely on univariate P-values for variable selection risks discarding key information that, while individually lacking statistical significance, may have joint predictive value. Therefore, while ensuring balance in age and sex between the two groups, we retained the maximum possible sample size to meet modeling requirements. Given that random forest (RF) and XGBoost, as ensemble tree models, incorporate multiple anti-overfitting mechanisms such as bootstrap sampling, tree pruning, and regularization, they can automatically quantify feature importance and attenuate the impact of noisy variables. Consequently, inputting high-dimensional variables does not substantially compromise the generalization ability of these models, and all 35 candidate variables were included in their construction. In contrast, logistic regression (LR) and support vector machines (SVM) lack native adaptive noise-reduction mechanisms, and the direct inclusion of all variables tends to amplify noise and induce overfitting. Therefore, for these two models, univariate feature prescreening was performed prior to modeling.

The results of 10-fold cross-validation for the four machine learning models are presented (Table 2 and Figure 3). Random Forest achieved the best overall performance, with a mean AUC of 0.9056 ± 0.0697, accuracy of 0.8000 ± 0.0987, sensitivity of 0.8111 ± 0.1892, specificity of 0.7889 ± 0.1105, and Kappa value of 0.6000 ± 0.1974. Notably, the Random Forest also exhibited the smallest standard deviations for both AUC and Kappa across folds, indicating the most robust generalization performance.

Following Random Forest, XGBoost ranked second, with a mean AUC of 0.8617 ± 0.1277 and sensitivity comparable to Random Forest (0.8111 ± 0.1576). However, its higher standard deviations (AUC SD: 0.1277, Kappa SD: 0.3045) indicated greater performance variability across different data partitions. Logistic Regression demonstrated moderate but highly stable performance, with the lowest AUC standard deviation (0.0476) among all models. Its AUC was 0.8272 ± 0.0476, but its sensitivity (0.7444 ± 0.0915) was lower than that of the two ensemble models, suggesting a higher risk of missed diagnosis. Finally, SVM exhibited the weakest overall performance, with a mean AUC of 0.7901 ± 0.1744, the lowest sensitivity (0.6889 ± 0.1554), and the largest standard deviations across all metrics, indicating insufficient generalization capacity. These comparative results are visually confirmed by the ROC curves (Figure 3): the Random Forest curve (Figure 3b) consistently demonstrated the best discriminative ability across all threshold levels, followed by XGBoost (Figure 3d), Logistic Regression (Figure 3a), and SVM (Figure 3c).

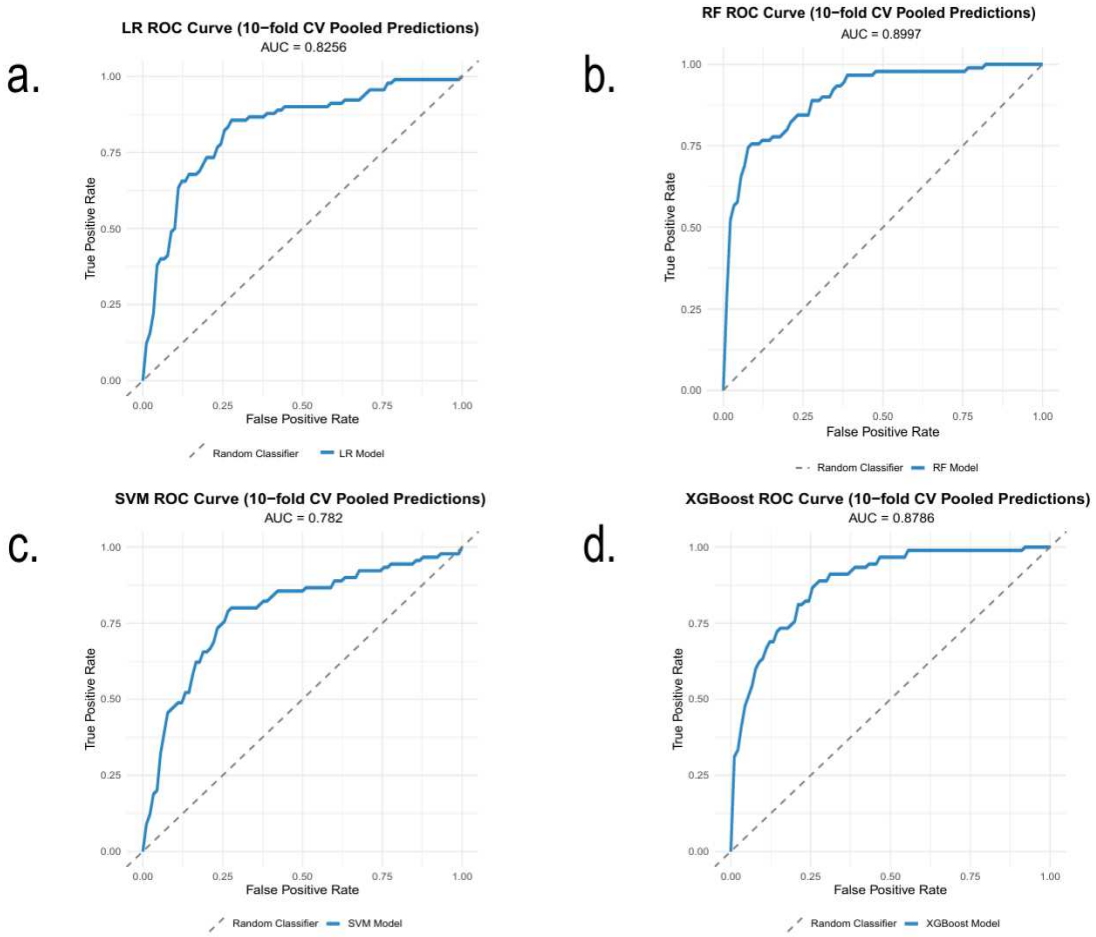


Table 2. Cross-validation performance of four models

Variables	RF	XGBoost	LR	SVM
AUC	0.9056±0.0697	0.8617±0.1277	0.8272±0.0476	0.7901±0.1744
Accuracy	0.8000±0.0987	0.7889±0.1523	0.7556±0.0597	0.7278±0.1321
Sensitivity	0.8111±0.1892	0.8111±0.1576	0.7444±0.0915	0.6889±0.1554
Specificity	0.7889±0.1105	0.7667±0.1693	0.7667±0.0973	0.7667±0.1693
Precision	0.7986±0.1038	0.7807±0.1361	0.7666±0.0714	0.7611±0.1592
F1	0.7939±0.1200	0.7934±0.1411	0.7517±0.0636	0.7158±0.1319
Kappa	0.6000±0.1974	0.5778±0.3045	0.5111±0.1194	0.4556±0.2642

Confusion matrix analysis (Supplementary Figure 2) revealed that RF and XGBoost achieved the best performance, with 17 false negatives each—a missed diagnosis rate of only 18.89%—thereby minimizing the risk of missed brucellosis diagnoses, which is particularly advantageous for primary care screening settings. Among the two, RF exhibited the best overall balance, with 19 false positives (misdiagnosis rate of 21.11%), slightly lower than XGBoost‘s 21 false positives (23.33%), achieving an optimal trade-off between the risks of missed diagnosis and misdiagnosis.

SHAP interpretability analysis of Random Forest and XGBoost

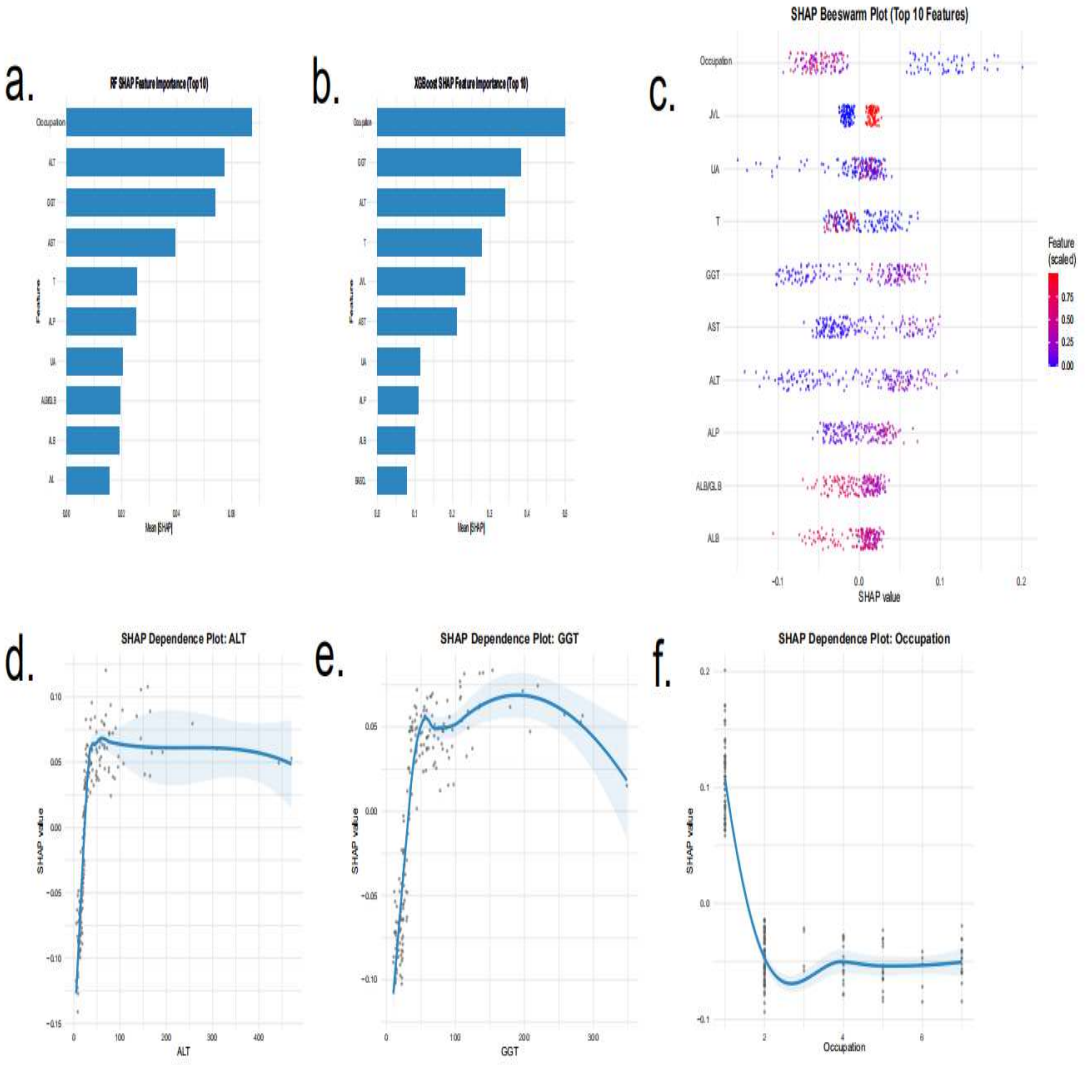
Given the comparable discriminative performance of the random forest (RF) and XGBoost models, we employed SHapley Additive exPlanations (SHAP) for interpretability analysis of both models. Global feature importance ranking based on mean absolute SHAP values revealed that occupation, alanine aminotransferase (ALT), and gamma-glutamyl transferase (GGT) were the top three core discriminative features for the RF model (Figure 4a); the XGBoost model showed an identical feature importance ranking (Figure 4b). These results confirm that these three variables are the most critical predictors for distinguishing between brucellosis and tuberculosis, a finding that is highly robust across the two ensemble learning algorithms.

The SHAP beeswarm plot for the RF model (Figure 4c) displays feature values ranging from low (blue) to high (red). On the horizontal axis, positive SHAP values

drive the model prediction toward brucellosis, whereas negative values drive it toward tuberculosis. Farmer/herder occupation (Occupation = 1, low value, blue), high ALT values (red), and high GGT values (red) all correspond to positive SHAP values, significantly shifting the prediction toward brucellosis; non-farmer/herder occupations (Occupation ≥ 2, high values, red), low ALT values (blue), and low GGT values (blue) correspond to negative SHAP values, shifting the prediction toward tuberculosis. These findings are consistent with the clinical, pathophysiological, and epidemiological features of the two diseases.

Partial dependence plots further quantified the marginal effects of the core features (Figure 4d – f). ALT showed a positive correlation with SHAP values; when ALT exceeded 40 U/L, SHAP values shifted from negative to positive, and the contribution to brucellosis prediction increased continuously as ALT rose. GGT also exhibited a positive correlation with SHAP values; when GGT exceeded 80 U/L, SHAP values became positive and increased progressively with higher GGT levels. For the multi-category variable occupation, the farmer/herder category (Occupation = 1) corresponded to positive SHAP values, whereas all other non-farmer/herder occupations were associated with negative SHAP values; moreover, larger occupation codes (indicating stronger non-farmer/herder status) were associated with lower SHAP values. These results clearly demonstrate the independent and core contribution of occupational exposure to differential diagnosis. Collectively, these findings provide intuitive quantitative references for clinical differential diagnosis and confirm that the

model's decision logic aligns with clinical knowledge, demonstrating strong clinical interpretability.



Optimal model selection

Because previous results showed that the performance of RF and XGBoost was significantly superior to that of the other two models, and both models selected the same set of variables, they were considered candidate models for further comparison using paired tests. Based on the paired prediction results across all samples in ten-fold cross-validation, McNemar's test was performed to compare the classification

accuracies of RF and XGBoost. The results showed no statistically significant difference between the two models (P = 0.7905) (Supplementary Table 3). However, RF had a higher mean cross-validation AUC than XGBoost (0.9056 vs. 0.8617) and exhibited smaller fold-to-fold fluctuations (0.0697 vs. 0.1277), indicating that RF had better discriminative ability and superior generalization stability. Furthermore, RF outperformed XGBoost in terms of both the mean cross-validation values and stability for accuracy, specificity, and the Kappa coefficient. Considering discriminative ability, classification stability, and generalization robustness, RF was ultimately selected as the optimal model.

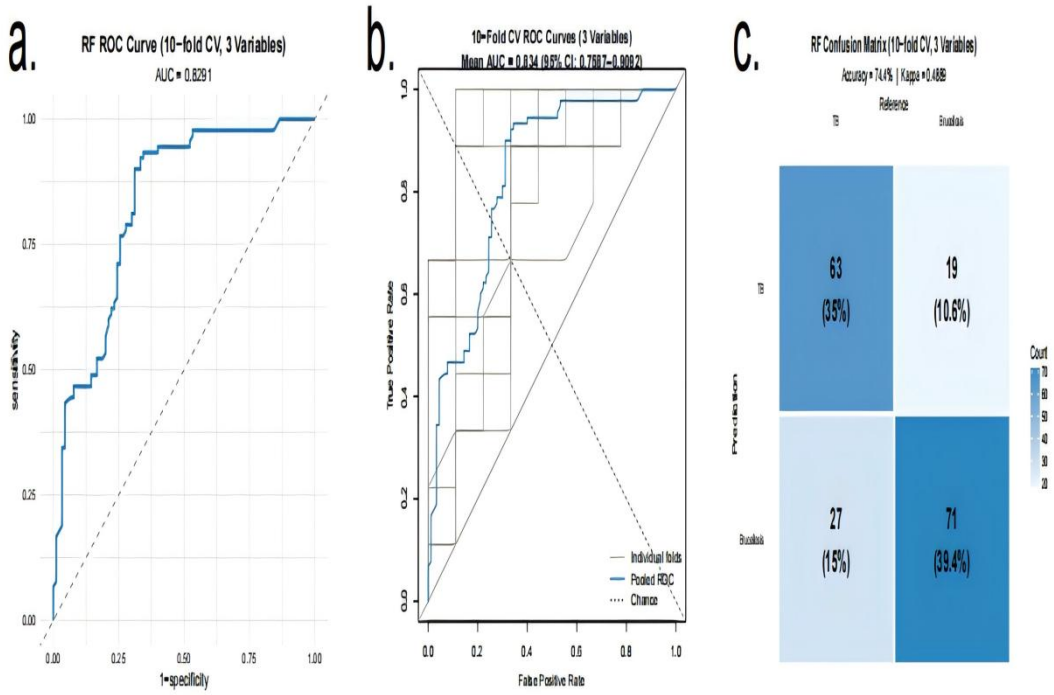
In summary, SHAP analysis not only identified the three core predictive indicators—ALT, GGT, and occupation—but also elucidated the specific mode of action of each indicator on the model, thereby substantially improving the interpretability of the differential diagnostic model. These findings are highly consistent with the pathophysiological characteristics and epidemiological patterns of the two diseases, further supporting the clinical applicability of the model. Given the RF model's comprehensive performance, this study aims to construct a more parsimonious RF model based on the three core features described above.

Simplified RF Model

To assess multicollinearity among the top ten features ranked by SHAP importance, we performed correlation and hierarchical clustering analyses (Supplementary Figure 3). The results showed no strong correlations ($|r| > 0.7$) among these ten features. Among

427 the three variables intended for inclusion in the parsimonious model—occupation, ALT,
428 and GGT—the correlation coefficients of occupation with ALT and with GGT were
429 0.08 and 0.11, respectively, indicating near-zero correlations; the correlation between
430 ALT and GGT was 0.62, representing a moderate positive correlation that did not
431 exceed the multicollinearity warning threshold. These findings indicate that
432 constructing a simplified random forest (RF) model based on these three core variables
433 can achieve maximal simplification while preserving discriminative performance, with
434 solid statistical and clinical rationale.

435 The ten-fold cross-validation results of the three-variable simplified RF model are
436 presented in Figure 5. The AUC for a single fold was 0.8291 (Figure 5a); the mean
437 AUC across ten-fold cross-validation was 0.834 (95% CI: 0.7587–0.9092) (Figure 5b).
438 The confusion matrix showed an overall accuracy of 74% and a Kappa coefficient of
439 0.488 (Figure 5c). The standard deviation of AUC across folds was 0.1052, lower than
440 that of XGBoost (0.1277) and SVM (0.1744), indicating that the simplified model had
441 smaller variability and stronger generalization robustness (Table 3).



442

443 **Table 3. 10-fold cross-validation performance of the 3-variable simplified**

444 **RF model**

	AUC	Accuracy	Sensitivity	Specificity	Precision	F1	Kappa
Cross-validation mean ± Standard deviation	0.834 ±0.1052	0.7444 ±0.1021	0.7889 ±0.1223	0.7 ±0.1486	0.7321 ±0.1104	0.755 ±0.0992	0.4889 ±0.2042

445

446 **Table 4. Summary of performance retention rates between simplified**

447 **models and full-feature models**

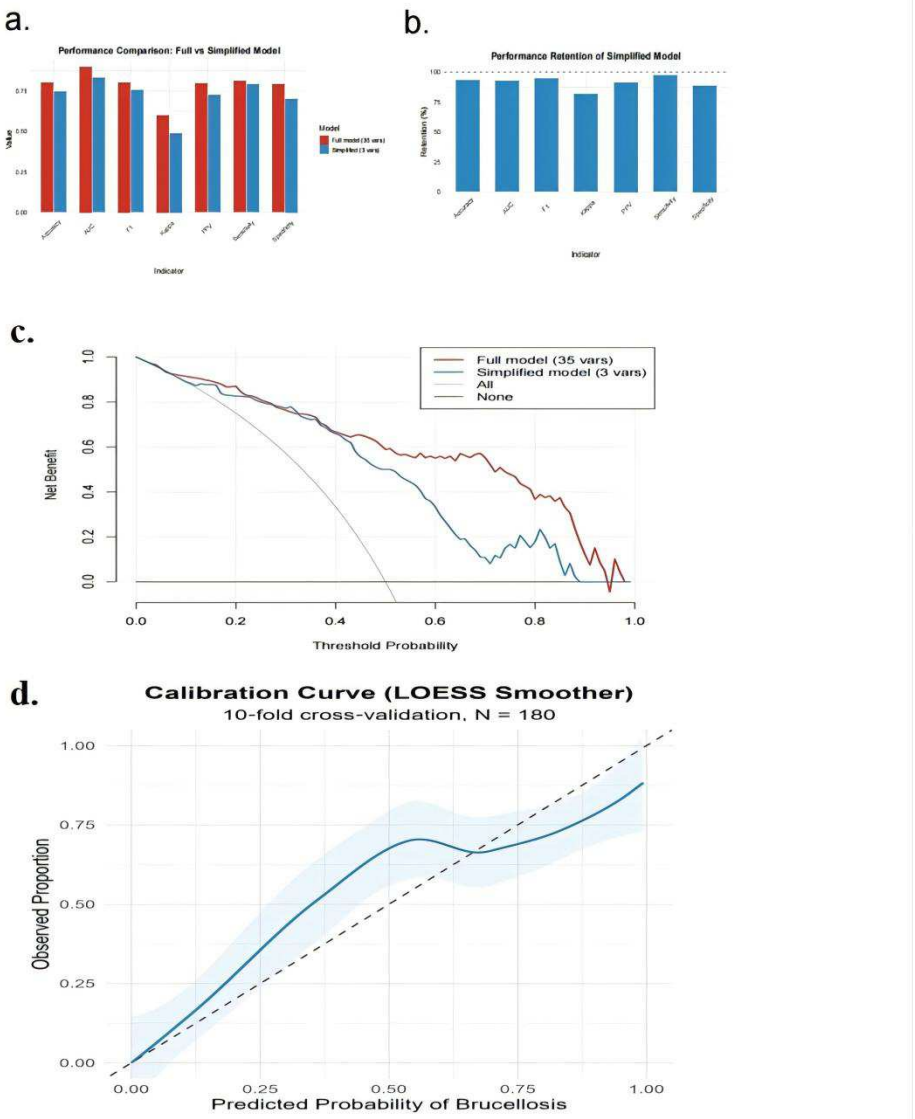
Indicator	Full	Simple	Retention
AUC	0.9056	0.834	92.09%
Accuracy	0.800	0.7444	93.05%
Sensitivity	0.8111	0.7889	97.26%
Specificity	0.7889	0.7000	88.73%
PPV	0.7986	0.7321	91.67%
F1	0.7939	0.755	95.10%
Kappa	0.600	0.4889	81.48%

448

449 Compared with the full-feature RF model, the performance retention rates
450 (simplified model / full-feature model × 100%) of the simplified RF model were as
451 follows: sensitivity 97.26%, AUC 92.09%, accuracy 93.06%, F1 score 95.1%;
452 specificity decreased from 0.7889 to 0.7000 (retention rate 88.73%), and Kappa
453 decreased from 0.6000 to 0.4889 (retention rate 81.48%) (Figure 6a, 6b; Table 4).
454 McNemar‘s test showed that the full-feature model was correct while the simplified
455 model was incorrect in 18 cases, and the simplified model was correct while the
456 full-feature model was incorrect in 8 cases; the difference was not statistically
457 significant ($\chi^2 = 3.1154$, $P = 0.0776$), although a trend toward lower accuracy was
458 observed for the simplified model (Supplementary Table 4).

459 Decision curve analysis showed that within the clinically reasonable probability
460 range of 0–0.4, the net benefit curves for the simplified and full-feature models were
461 essentially overlapping (Figure 6c). At the commonly used threshold probability of 0.2,

462 the net benefit of the full-feature model was 0.382, and that of the simplified model was
463 0.376, a difference of only 0.006 (Figure 6c). Calibration curves were plotted using
464 LOESS (locally weighted scatterplot smoothing) based on ten-fold cross-validation
465 results to assess the calibration performance of the simplified model (Figure 6d). The
466 results showed that the fitted curve of the simplified model aligned well with the ideal
467 calibration line (predicted probability = actual probability). Within the most widely
468 used decision threshold range of 0.2–0.4, the model‘s predicted probabilities were
469 highly consistent with the actual proportions of brucellosis cases; only a slight
470 overestimation was observed in the low-probability range (0–0.25), and a slight
471 underestimation was observed in the high-probability range (0.6–1.0), both of which
472 were within clinically acceptable limits, confirming that the simplified model has good
473 calibration performance (Figure 6d).



In summary, the simplified model maintained retention rates exceeding 90% for sensitivity, AUC, accuracy, and F1 score, with a specificity retention rate of 88.73% and a Kappa retention rate of 81.48%. This simplified model can substantially reduce the cost of feature variable collection without significantly compromising core diagnostic performance, thereby enhancing its suitability for promotion in primary care settings.

Temporal external validation of the simplified RF model

To evaluate the temporal generalizability of the simplified model, we collected an independent temporal external test set comprising 42 cases (21 with brucellosis and 21 with tuberculosis), including only the three variables (GGT, ALT, and occupation). The collection period of this test set did not overlap with that of the primary dataset (Table 5). The final simplified Random Forest model, trained on the entire primary dataset of 180 cases, was applied to make independent predictions on this test set using the clinically conventional threshold of 0.5 for classification. Importantly, none of the test set samples were involved in model training or parameter optimisation.

Temporal external validation of the simplified model was performed using independent samples from different time points (Table 5). The model achieved an AUC of 0.8265 (95% CI: 0.6956–0.9574), an accuracy of 78.57%, a sensitivity of 85.71%, a specificity of 71.43%, a precision of 75.00%, an F1 score of 0.800, and a Kappa coefficient of 0.5714, indicating moderate agreement between model predictions and true labels (Figure 7a, 7b).

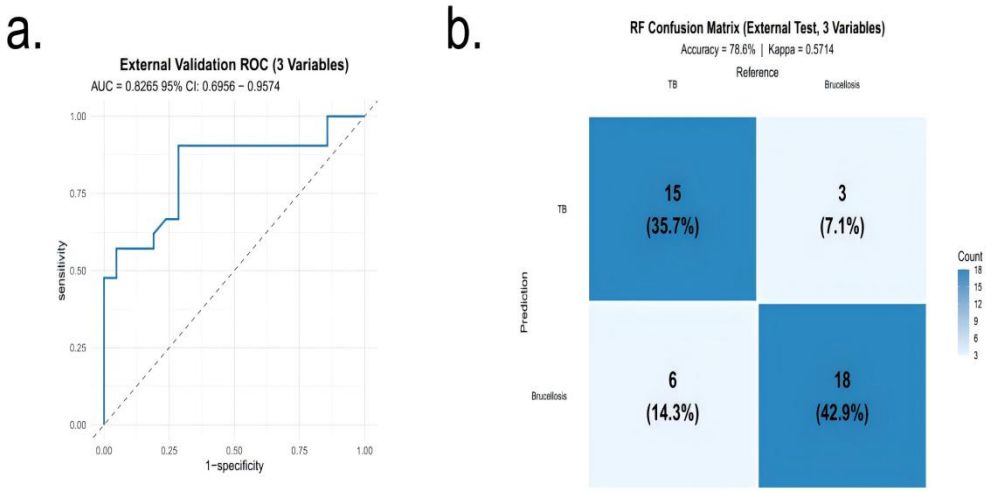


Table 5. Performance of the simplified model in the external validation cohort

Metric	Value
AUC	0.8265
Accuracy	0.7857
Sensitivity	0.8571
Specificity	0.7143
Precision	0.7500
F1	0.8000
Kappa	0.5714

External validation results were highly consistent with cross-validation findings (AUC: 0.8340 vs. 0.8265; sensitivity: 78.89% vs. 85.71%; specificity: 70.00% vs. 71.43%), confirming that the simplified model maintained stable discriminative capacity and sensitivity-prioritized diagnostic performance across samples from different time periods, with good generalization robustness. The sensitivity in the external test set (85.71%) was slightly higher than that in internal cross-validation (78.89%), which may be attributed to more typical cases in the external sample and statistical fluctuations due to the small sample size. Overall, the model retained its sensitivity-prioritized diagnostic characteristics, meeting the clinical requirements for primary care screening.

These findings confirm that the simplified model maintains stable discriminative performance in independent samples from different time points, with good generalization capacity. Overall, the model, which is highly sensitive, is suitable for clinical application in primary care screening settings.

Discussion

Brucellosis and tuberculosis are highly detrimental chronic infectious diseases, causing increased disability and mortality. Both diseases often present with similar symptoms, such as fever, fatigue, and weight loss, complicating their differentiation. They also impose a substantial economic burden on healthcare systems. This challenge is especially pronounced in pastoral regions such as Xinjiang, China[16]. Although brucellosis has a lower overall incidence than tuberculosis, diagnosis and treatment are often delayed. These delays can readily lead to chronicity and irreversible complications, including osteoarticular damage and persistent hepatic injury. Such complications severely compromise patients' quality of life. Tuberculosis, a globally high-burden infectious disease, remains endemic in Xinjiang. The considerable overlap in clinical presentations makes differential diagnosis difficult in co-endemic areas[2, 17]. Diagnostic delay drives poor prognosis and increases healthcare costs. Integrating explainable machine learning algorithms with routine admission blood biomarkers enhances diagnostic accuracy. It also enables visualization of the model’s decision-making logic, which is essential for fostering primary care clinicians’ acceptance of intelligent adjunctive tools. Collectively, our findings confirm that explainable machine learning can be a rapid, accessible, and low-cost diagnostic aid for differentiating brucellosis from tuberculosis, especially in resource-limited settings. limited primary care settings.

This study implemented distinct feature inclusion strategies for four machine learning algorithms, aligning each approach with the specific characteristics of the

539 respective algorithm. Univariate analysis captures only the independent discriminatory
540 capacity of each indicator; exclusive reliance on P-value thresholds for variable
541 selection may eliminate predictors whose significance arises from interaction effects,
542 potentially introducing selection bias[18, 19]. Random Forest (RF) and XGBoost, as
543 ensemble tree-based methods, mitigate overfitting through mechanisms such as random
544 subsampling, tree pruning, and integrated regularization. Consequently, these
545 algorithms can utilize the full feature set without prior dimensionality reduction[20,
546 21].In contrast, Logistic Regression (LR) and Support Vector Machine (SVM) lack
547 adaptive noise filtering and are more susceptible to noise when trained on high-
548 dimensional data, necessitating stringent variable selection. Within this comparative
549 framework, RF and XGBoost used the full feature set, whereas LR and SVM underwent
550 rigorous feature screening. This approach ensures that each algorithm's operational
551 logic is appropriately matched to its feature selection method, differentiates between
552 tree-based and linear models, and facilitates optimal identification of complex variable
553 interactions relevant to distinguishing between the two infections.

554 A systematic comparison of four machine learning algorithms identified Random
555 Forest as the optimal modeling approach. Integration of SHAP-based interpretability
556 analysis revealed occupation, alanine aminotransferase (ALT), and γ-glutamyl
557 transferase (GGT) as the three primary discriminatory features, leading to the
558 construction of a lightweight Random Forest-based auxiliary diagnostic model. Internal
559 cross-validation and temporally independent external validation demonstrated that the
560 model delivers stable performance, with high sensitivity, favorable interpretability, and

considerable clinical net benefit. These results underscore its potential as a clinical decision support tool. SHAP dependence plots indicated that ALT and GGT levels were significantly positively correlated with the model-predicted probability of brucellosis, and that a background in farming or herding also increased the likelihood of a brucellosis prediction. These findings align with the distinct pathophysiological profiles of the two diseases: Brucella exhibits tropism for the reticuloendothelial system, frequently invading the liver and inducing hepatic injury[22, 23]，which is accompanied by elevated peripheral GGT and ALT levels[6, 24, 25].Farming and herding populations are at substantially higher risk of infection due to routine occupational exposure[6]. In contrast, tuberculosis is characterized primarily by fever and relatively higher serum albumin levels. Based on these insights, a parsimonious Random Forest model was developed using only occupation, ALT, and GGT. Ten-fold internal cross-validation showed that the simplified model retained over 90% of the full model's core performance metrics. feature model. Decision curve analysis indicated that, within clinically relevant threshold ranges, the net benefit curves of the simplified and full? feature models were largely overlapping. In summary, multi-Algorithm comparison and SHAP-based explainability analysis demonstrate that a simplified Random Forest model, built on routine liver function indicators and occupational information, can serve as a reliable auxiliary tool for differentiating brucellosis from tuberculosis.

Model generalizability is a fundamental criterion for assessing clinical utility. In this study, the simplified Random Forest model was externally validated using a temporally independent cohort. The findings indicate that the model maintained stable

583 discriminative performance in the independent sample, with area under the curve (AUC)
584 and sensitivity values closely matching those from internal cross-validation. The
585 external validation set demonstrated slightly higher sensitivity than internal cross-
586 validation. This difference may result from more typical clinical presentations in the
587 external cohort or from random variation due to the relatively small sample size. The
588 model's primary strength is its high sensitivity, which aligns with the clinical priority
589 in primary care screening to minimize missed diagnoses of brucellosis. Although multi-
590 center external validation has not yet been conducted, the training and temporal external
591 cohorts were similar in demographic and clinical baseline characteristics, and no
592 significant decline in discriminative performance was observed. This suggests that the
593 model possesses reasonable transportability within the local healthcare system. The
594 study design addresses a central challenge in implementing clinical artificial
595 intelligence: integrating high-performance machine learning algorithms with clinical
596 decision support. Additionally, cross-referencing model predictions with established
597 pathophysiological mechanisms enhances the clinical credibility of the diagnostic
598 output.

599 This study identified gamma-glutamyl transferase (GGT) and alanine
600 aminotransferase (ALT) as the principal biochemical markers for differentiating
601 brucellosis from tuberculosis, a finding attributable to the distinct pathogenic
602 mechanisms of the two diseases. Brucella preferentially colonizes the
603 reticuloendothelial system (RES), forming non-caseating necrotizing granulomas
604 within the liver and inducing diffuse hepatocellular injury, which results in substantial

ALT release into the peripheral circulation[25-27].As the largest organ of the RES, the liver is crucial for maintaining immune function, yet it is also susceptible to various pathological insults. Brucellosis is a multisystemic infectious disease; the bacterium can invade hepatic macrophages and cause hepatocyte damage, which may in turn lead to abnormal elevations in liver enzyme levels[27].Previous studies have confirmed that the liver is among the most frequently affected organs in patients with active brucellosis, and that mild-to-moderate elevation of serum transaminases is a characteristic manifestation of hepatic dysfunction. Therefore, ALT, rather than AST, more specifically reflects hepatocellular membrane injury and defines the acute-phase hepatic involvement in brucellosis, accounting for the emergence of ALT (not AST) as a core predictive variable in this study. Liver colonization involves Brucella binding to hepatic sinusoidal endothelial cells via the adhesin BtaE, which recognizes hepatocyte-specific glycoproteins, thereby facilitating bacterial survival within the RES. The coordinated expression of these adhesins enables Brucella to establish infection in diverse tissues, ensuring long-term persistence and systemic dissemination[28].

Meanwhile, *Brucella* infection induces pronounced oxidative stress, leading to a significant depletion of reduced glutathione (GSH). In response to this metabolic disturbance, the body compensates by upregulating GGT expression, resulting in elevated serum GGT levels[24, 29].Animal studies have demonstrated that hepatic antioxidant enzyme activity is markedly reduced in Brucella-infected goats, accompanied by a sharp increase in serum ALT and GGT[29]. This pattern is consistently observed in the present cohort data. Recent evidence suggests that Brucella

627 infection activates host adaptive immune responses while disrupting innate immune
628 homeostasis, and these immunological changes are closely associated with hepatic
629 injury[6, 13, 30].In contrast, *Mycobacterium tuberculosis* primarily colonizes the
630 respiratory tract and causes less extensive and persistent hepatic involvement than
631 *Brucella*, resulting in more limited hepatocellular damage. As a result, patients with
632 brucellosis display significantly higher peripheral ALT and GGT levels than those with
633 tuberculosis, a baseline intergroup difference that is consistent with the SHAP analysis
634 in this study. Overall, GGT and ALT not only reflect the extent of liver injury caused
635 by Brucella infection but also serve as effective discriminatory markers between these
636 two diseases.

637 This study used multiple machine learning algorithms to systematically evaluate
638 the discriminatory performance of clinical biochemical markers. ALT and GGT
639 showed greater discriminatory power than conventional inflammatory markers in
640 differentiating brucellosis from tuberculosis. These findings expand the understanding
641 of oxidative stress and hepatocellular injury caused by *Brucella* infection. Our
642 analytical approach intuitively reflects mechanistic differences between Brucella and
643 *Mycobacterium tuberculosis* infections. It also shows that, compared to M. tuberculosis,
644 Brucella creates a more pronounced disturbance of host hepatic function. This study is
645 the first to develop a machine-learning-based diagnostic tool for discriminating
646 between brucellosis and tuberculosis using routine biochemical indicators and
647 occupational exposure information.

Although the model shows favorable statistical performance, several limitations in clinical validation exist. First, this study's propensity score matching only adjusted for age and sex. Unmeasured confounders, such as alcohol use, non-alcoholic fatty liver disease, and viral hepatitis, may still affect liver function test results. Second, the external validation cohort was from different periods at the same institution. Cross-center validation using samples from other regions or facility types was not done, which limits the model's generalizability. Third, all patients enrolled were definitively etiologically positive. They may have stronger immune responses than suspected cases without microbiological evidence, potentially introducing selection bias. Future studies should include clinically diagnosed cases to test the model's ability to identify early or atypical infections. Fourth, because this study was retrospective and single-center, it could not evaluate the model's real-world utility. Impacts on diagnostic turnaround time, misdiagnosis, and missed diagnosis rates, or longer-term patient outcomes such as brucellosis chronicity or drug-resistant tuberculosis incidence, were not assessed. Prospective interventional studies are needed to determine whether integrating this model into primary care can shorten diagnostic time and improve patient outcomes.

Conclusions

In this study, we systematically compared four machine learning algorithms and identified random forest (RF) as the optimal model (cross-validated AUC = 0.9056 ± 0.0697), outperforming XGBoost and linear models in both discriminative ability and stability. SHAP analysis then identified GGT, ALT, and occupation as core predictive

variables and quantified their directional effects: ALT > 40 U/L, GGT > 80 U/L, and farmer/herder occupation all drove the model's prediction of brucellosis. This dose-response relationship is pathophysiologically consistent with Brucella-induced liver injury and occupational exposure risk, thereby endowing the model with clinical interpretability. Building on these findings, a simplified RF model based on these three variables retained 92.09% of the full-model AUC and 97.26% of its sensitivity. McNemar's test (P = 0.0776) and decision curve analysis further confirmed that the simplified model was not statistically different from the full model and provided comparable net clinical benefit. Temporal external validation yielded an AUC of 0.8265 (95% CI: 0.6956–0.9574) and a sensitivity of 85.71%, confirming its generalizability across time. Collectively, this study presents a complete chain of evidence through algorithm comparison, feature elucidation, model simplification, and external validation, providing a low-cost, high-sensitivity, and interpretable auxiliary diagnostic tool for primary healthcare settings in co-endemic regions, relying solely on routine liver function tests and occupational history.

Abbreviations

ALT: Alanine aminotransferase;

GGT: γ -glutamyl transferase;

PSM: Propensity score matching

SHAP: SHapley Additive exPlanations;

SVM: Support vector machine;

692 TB: Tuberculosis;

693 XGBoost: Extreme gradient boosting

694 RF:Random Forest

695

696 **Supplementary Information**

697 Additional details are provided in the Supplementary Materials, including
698 supplementary figures and tables.

699

700 **Acknowledgements**

701 We acknowledge the Beckman Coulter DxAI platform
702 (https://www.xsmartanalysis.com/beckman/login/) for providing technical support to
703 complete part of the data analysis in this study.

704

705 **Authors' contributions**

706 SW: conceptualization, methodology, writing – original draft; ZJP: methodology,
707 validation, writing – original draft; XRX: formal analysis, methodology, writing –
708 original draft ; JJD: investigation, data collection; JJM: investigation, data collection;
709 ZY: visualization, validation; XX: investigation; GLD: validation; LYM: funding
710 acquisition, writing – review & editing, supervision; XMM: resources, writing – review
711 & editing.

712

Funding

The authors declare that financial support for the research and publication of this paper was provided by the Open Fund of the State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia (SKL-HIDCA-2024-10), the Central Government Guides Local Science and Technology Development Special Fund (ZYYD2024CG06), and the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2025D01C211).

Data availability

All data generated or analyzed during this study are included in this article. For additional requests, please contact the corresponding author.

Declarations

Ethics approval and consent to participate

The Ethics Committee of the Affiliated Hospital of Traditional Chinese Medicine, Xinjiang Medical University (approval No. 2026XE0115), approved this retrospective study involving human participants. The study adhered to local laws, regulations, and institutional requirements. The Ethics Committee granted a waiver of informed consent in accordance with local policies. All patient data were anonymized for analysis.

Consent for publication

All patient data analyzed in this retrospective study were fully de-identified and anonymized prior to use. No identifiable individual information, including names,

735 medical record numbers, images, or personal identifiers, is presented in this manuscript,
736 tables, or figures. The Ethics Committee of the Affiliated Hospital of Traditional
737 Chinese Medicine, Xinjiang Medical University, formally waived the requirement for
738 individual written consent for publication, given the completely anonymized and non-
739 identifiable nature of the retrospective clinical data. Accordingly, patient consent for
740 publication was not required.

741

742 **Competing interests**

743 The authors declare no commercial or financial relationships that could be seen as a
744 conflict of interest.

745

746 **Author details**

747 1. State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence
748 Diseases in Central Asia, Clinical Laboratory Center, Tumor Hospital Affiliated to
749 Xinjiang Medical University, Urumqi 830000, P. R. China
750 2. Clinical Laboratory Center, Xinjiang Medical University Affiliated Traditional
751 Chinese Medicine Hospital, Urumqi 830000, P. R. China

752

753 **Table 1. Comparison of baseline characteristics between patients with**
754 **brucellosis and tuberculosis**

Variable	Overall N=180[1]	TB N=90[1]	Brucellosis N=90[1]	*P*-value[2]

Age	56.000(52.000,66.000)	56.000(52.000,67.000)	56.000(52.000,63.000)	0.348
Gender	125(69.4%)	61(67.8%)	64(71.1%)	0.627
Drink	23(12.8%)	12(13.3%)	11(12.2%)	0.823
Smoking	40(22.2%)	25(27.8%)	15(16.7%)	0.073
Temperature	0.085(0.020,0.235)	0.118(0.070,0.290)	0.050(0.010,0.130)	<0.001
Joint and Vertebral Lesions	85(47.2%)	29(32.2%)	56(62.2%)	<0.001
Hypertension	51(28.3%)	25(27.8%)	26(28.9%)	0.869
Diabetes mellitus	27(15.0%)	15(16.7%)	12(13.3%)	0.531
Occupations				<0.001
farmers and herdsmen	56(31.1%)	8(8.89%)	48(53.3%)	
retirees	69(38.3%)	45(50.0%)	24(26.7%)	
state servants	4(2.22%)	2(2.22%)	2(2.22%)	
Self-employed households	15(8.33%)	11(12.2%)	4(4.44%)	
staff	16(8.89%)	12(13.3%)	4(4.44%)	
professional and technical personnel	5(2.78%)	3(3.33%)	2 (2.22%)	
unemployed/others	15 (8.33%)	9(10.0%)	6(6.67%)	
White blood cell count (WBC)	0.030(0.020,0.070)	0.040(0.020,0.090)	0.030(0.020,0.050)	0.078

Absolute neutrophil count (ANC)	120.500(91.500,135.500)	123.500(89.000,137.000)	119.000(92.000,133.300)	0.554
Absolute lymphocyte count (LYM)	4.285(3.445,4.676)	4.350(2.970,4.670)	4.229(3.600,4.681)	0.767
Absolute monocyte count (MO)	208.000(102.000,262.500)	214.500(132.000,273.000)	197.500(84.000,247.000)	0.204
Absolute eosinophil count (EOS)	23.065(1.840,46.415)	21.155(0.940,41.387)	26.970(2.440,53.900)	0.412
Absolute basophil count (BASO)	0.097(0.040,0.328)	0.176(0.040,0.419)	0.070(0.040,0.190)	0.014
Platelet count (PLT)	0.080(0.060,0.135)	0.080(0.069,0.160)	0.080(0.060,0.130)	
Hemoglobin (Hb)	73.130(61.300,95.945)	73.010(61.190,99.390)	73.400(62.300,86.000)	0.655
Red blood cell (RBC)	4.745(4.055,6.235)	4.870(4.100,6.130)	4.690(4.010,6.350)	0.564
Procalcitonin (PCT)	4.710(4.045,5.950)	4.525(3.900,6.120)	4.895(4.320,5.830)	0.237

High-sensitivity C-reactive protein (hsCRP)	225.650(139.250,293.900)	247.850(139.900,309.700)	203.600(138.400,275.100)	0.105
Lactate dehydrogenase (LDH)	211.650(179.650,288.135)	188.060(172.600,235.300)	247.825(192.000,369.200)	<0.001
Creatine kinase MB isoenzyme (CKMB)	10.067(8.135,13.440)	10.342(8.540,13.820)	9.905(7.680,13.200)	0.443
Creatine kinase (CK)	54.000(34.250,78.475)	57.300(39.500,79.200)	50.684(28.500,74.300)	0.126
Glucose (GLU)	0.640 (0.470, 1.045)	0.630(0.500,0.820)	0.640(0.420,1.180)	0.840
Uric acid (UA)	1.750(1.219,3.005)	1.652(1.030,2.470)	1.970(1.490,3.380)	0.024
Blood urea nitrogen (BUN)	36.400(36.100,36.600)	36.300(36.100,36.600)	36.400(36.000,36.600)	0.569
Creatinine (CREA)	7.190 (5.545, 11.645)	7.215(6.020,10.970)	7.171(5.330,12.090)	0.785
BUN/CREA	3.745 (1.680, 4.869)	3.950(2.230,5.000)	3.245(1.640,4.804)	0.395
Gamma-glutamyl transferase (GGT)	37.625(22.130,69.980)	23.650(17.560,43.900)	56.720(35.600,86.500)	<0.001
Total protein (TP)	67.100(63.010,71.885)	66.800(63.400,72.000)	67.200(62.400,71.700)	0.870
Albumin (ALB)	34.500 (31.750, 38.300)	36.250(32.740,39.800)	33.860(30.500,36.380)	0.001
Globulin (GLB)	31.450 (28.825, 35.835)	30.100(28.300,33.900)	33.170(29.700,37.000)	0.002

ALB/GLB	1.070(0.910,1.295)	1.200(1.000,1.390)	1.020(0.870,1.170)	<0.001
Alkaline phosphatase (ALP)	94.200(73.450,125.450)	83.750(70.100,102.100)	106.750(87.210,154.000)	<0.001
Aspartate aminotransferase (AST)	26.150(18.250,47.205)	21.240(16.100,27.700)	38.530(24.800,62.980)	<0.001
Alanine aminotransferase (ALT)	23.665(15.130,49.950)	18.400(12.300,25.030)	37.925(23.000,69.400)	<0.001
AST/ALT	0.915 (0.705, 1.210)	0.870(0.680,1.080)	1.055(0.720,1.380)	0.013

[1] Median (IQR) or Frequency (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Figure Legends

Figure 1. Enrollment process flowchart. Eligible patients (N = 273) were matched 1:1 by propensity score matching (PSM) based on age and sex, resulting in 90 patients with brucellosis and 90 with tuberculosis. Candidate variables were screened via univariate analysis. Four machine learning models (LASSO-Firth logistic regression, RFE-SVM, Random Forest [RF], and XGBoost) were constructed and compared to select the optimal model. SHAP analysis guided the development of a simplified three-variable model, which was validated using internal cross-validation and an independent temporal external test set.

765

766 **Figure 2. Correlation matrix of routine clinical laboratory biomarkers.** Pairwise
767 Pearson correlation coefficients between 29 continuous laboratory parameters are
768 visualized as a heatmap. Each cell represents the correlation coefficient between the
769 corresponding row and column variables. The colour scale denotes the correlation
770 strength: pink (positive correlation) and green (negative correlation), with deeper
771 colours indicating stronger correlations (range: -1 to 1). Hierarchical clustering
772 dendrograms are shown on the top and left to illustrate the similarity of correlation
773 profiles across parameters. Absolute correlation coefficients for all variables were
774 below 0.8, indicating no severe multicollinearity.

775

776 **Figure 3. Ten-fold cross-validation ROC curves for the four machine learning**
777 **models.** (a) Logistic regression (LR), (b) Random Forest (RF), (c) support vector
778 machine (SVM), and (d) XGBoost. For each model, the thin grey lines represent the
779 ROC curves of individual folds, and the thick blue line represents the pooled ROC curve.
780 The mean area under the curve (AUC) values were 0.8272 for LR, 0.9056 for RF,
781 0.7901 for SVM, and 0.8617 for XGBoost. RF achieved the highest AUC and the
782 smallest cross-fold variability, indicating superior discriminative ability and
783 generalisation stability.

784

785 **Figure 4. SHAP analysis of the predictive model.** (a) The top ten features ranked by
786 mean absolute SHAP values in the Random Forest (RF) model. (b) The top ten features

787 in the XGBoost model. Both models identified occupation, alanine aminotransferase
788 (ALT), and γ-glutamyl transferase (GGT) as the three core discriminative features. (c)
789 SHAP beeswarm plot for the RF model showing the distribution and impact of the top
790 ten features on model output. Each point represents a single patient, coloured by the
791 scaled feature value (blue = low, red = high). Positive SHAP values drive prediction
792 toward brucellosis, and negative values toward tuberculosis. (d–f) SHAP dependence
793 plots for ALT (d), GGT (e), and occupation (f). The blue line represents the trend of
794 the average SHAP value, and the shaded area indicates the confidence interval.
795 Farmer/herder occupation, high ALT, and high GGT positively contributed to
796 brucellosis prediction.
797
798 **Figure 5. Performance evaluation of the simplified three-variable Random Forest**
799 **(RF) model.** (a) Receiver operating characteristic (ROC) curve of the simplified RF
800 model on the full dataset (N = 180), with an area under the curve (AUC) of 0.8291. (b)
801 Ten-fold cross-validation ROC curves. The blue line represents the pooled ROC curve
802 (mean AUC = 0.834, 95% CI: 0.7587–0.9092), and the grey lines denote the ROC
803 curves of individual folds. (c) Confusion matrix of the simplified RF model. The overall
804 accuracy was 74.4%, and the Kappa coefficient was 0.4889. TB, tuberculosis.
805
806 **Figure 6. Performance comparison and clinical utility of the full (35-variable) and**
807 **simplified (3-variable) models.** (a) Comparison of key performance metrics (accuracy,
808 AUC, F1 score, Kappa, positive predictive value [PPV], sensitivity, and specificity)

between the full and simplified models. (b) Performance retention of the simplified model relative to the full model, expressed as the percentage of each metric retained. Sensitivity, AUC, accuracy, and F1 score retention exceeded 90%. (c) Decision curve analysis (DCA) showing the net benefit of the full model (red line) and the simplified model (blue line) across different threshold probabilities. The grey lines represent the "treat all" and "treat none" strategies. The two curves largely overlap within the clinically reasonable threshold range (0–0.4). (d) Calibration curve (LOESS smoother) of the simplified model for brucellosis prediction, based on ten-fold cross-validation (N = 180). The blue line and shaded area represent the observed proportion and its 95% confidence interval, while the dashed line indicates perfect calibration. The fitted curve aligns well with the ideal calibration line.

Figure 7. External validation of the simplified three-variable Random Forest (RF) model on a temporally independent test set (n = 42). (a) Receiver operating characteristic (ROC) curve. The area under the curve (AUC) was 0.8265 (95% CI: 0.6956–0.9574). (b) Confusion matrix. The overall accuracy was 78.6%, sensitivity 85.7%, specificity 71.4%, and Kappa coefficient 0.5714, indicating moderate agreement between model predictions and true labels. TB, tuberculosis.

References

1. Tian Z, Wan L, Pei J, Li T, Wang X, Yuan P, et al. Brucellosis seroprevalence in cattle in China during 2014–2024: A systematic review and meta-analysis. Emerging Microbes & Infections. 2024;13:2417859. https://doi.org/10.1080/22221751.2024.2417859.

2. Raz KM, Haddad MB, Althomsons SP, Cowan L, Kammerer JS, Lam CK, et al. Large tuberculosis outbreaks — united states, 2017–2023. MMWR Morb Mortal Wkly. 2026;75:205–10. https://doi.org/10.15585/mmwr.mm7516a1.

3. Laine CG, Johnson VE, Scott HM, Arenas-Gamboa AM. Global estimate of human brucellosis incidence. Emerg Infect Dis. 2023;29. https://doi.org/10.3201/eid2909.230052.

4. Liu W, Ma J, Zhang J, Cao J, Hu X, Huang Y, et al. Identification and validation of serum metabolite biomarkers for endometrial cancer diagnosis. EMBO Mol Med. 2024;16:988–1003. https://doi.org/10.1038/s44321-024-00033-1.

5. Wang H, Wang YO, Yuan G, Li D, Chen Y, Wang YA, et al. Natural history of latent brucellosis and diagnostic performance of B-SPOT: A 24-month prospective cohort study based on large-scale screening in northern China. Emerging Microbes & Infections. 2026;15:2637283. https://doi.org/10.1080/22221751.2026.2637283.

6. Ghssein G, Ezzeddine Z, Tokajian S, Khoury CA, Kobeissy H, Ibrahim J-N, et al. Brucellosis: Bacteriology, pathogenesis, epidemiology and role of the metallophores in virulence: a review. Front Cell Infect Microbiol. 2025;15:1621230. https://doi.org/10.3389/fcimb.2025.1621230.

7. Nijiati M, Zhang M, Huang C, Chou X, Shen L, Ma H, et al. Multimodal-imaging-based interpretable deep learning framework for distinguishing brucella from tuberculosis spondylitis: A dual-center study. Diagnostics. 2025;15:2963. https://doi.org/10.3390/diagnostics15232963.

8. Javadi S, Razavi MR, Ghazvini K, Salehi M. Sensitivity of current laboratory methods for brucellosis diagnosis at endemic foci. BMC Microbiol. 2026;26:482. https://doi.org/10.1186/s12866-026-04842-x.

9. Wang R, He J, Li X, Shi Y, Duan H, Ding H. Comprehensive analysis of molecular immunological characteristics and potential biomarkers in brucellosis. Front Immunol. 2026;17:1614025. https://doi.org/10.3389/fimmu.2026.1614025.

10. Lin Z, Tang J, Chen Q, Mi H, Zhang W, Zhang J, et al. What are the clinical characteristics and outcomes of brucellar spondylitis in a nonendemic region of southern China? Clin Orthop Relat Res. 2026;2026 Apr 29. https://doi.org/10.1097/CORR.0000000000003968.

11. Han X, Zhang R, Xu B, Zhang W, Wang F, Yuan B, et al. Blood culture positivity and its clinical determinants among hospitalized male patients with newly diagnosed, laboratory-confirmed brucellosis: A hospital-based retrospective study. Front Microbiol. 2026;17:1754087. https://doi.org/10.3389/fmicb.2026.1754087.

12. Prodanuk M, Boillat-Blanco N, Fung A, Kain KC, Kazimoto T, Kitai I, et al. Host plasma protein biomarkers for tuberculosis disease screening in febrile adults in tanzania. J Infect Dis. 2026;jiag238. https://doi.org/10.1093/infdis/jiag238.

13. Wang R, Niu B, Zhang C, Wang Y, Zhang X, Tian H, et al. Machine learning-based prediction model for chronic brucellosis: A multi-feature approach using clinical and laboratory data. Front Cell Infect Microbiol. 2025;15:1700233. https://doi.org/10.3389/fcimb.2025.1700233.

14. Yan F, Chen X, Quan X, Wang L, Wei X, Zhu J. Association between the stress hyperglycemia ratio and 28-day all-cause mortality in critically ill patients with sepsis: A retrospective cohort study and predictive model establishment based on machine learning. Cardiovasc Diabetol. 2024;23:163. https://doi.org/10.1186/s12933-024-02265-4.

15. Theodosiou AA, Read RC. Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician. Journal of Infection. 2023;87:287–94. https://doi.org/10.1016/j.jinf.2023.07.006.

16. Wang Y, Yang S, Han B, Du X, Sun H, Du Y, et al. Single-cell landscape revealed immune characteristics associated with disease phases in brucellosis patients. iMeta. 2024;3:e226. https://doi.org/10.1002/imt2.226.

17. Roop RM, Barton IS, Hopersberger D, Martin DW. Uncovering the hidden credentials of *brucella* virulence. Microbiol Mol Biol Rev. 2021;85:e00021-19. https://doi.org/10.1128/MMBR.00021-19.

18. Hafermann L, Becher H, Herrmann C, Klein N, Heinze G, Rauch G. Statistical model building: Background “knowledge” based on inappropriate preselection causes misspecification. BMC Med Res Methodol. 2021;21:196. https://doi.org/10.1186/s12874-021-01373-z.

19. Ullmann T, Heinze G, Hafermann L, Schilhart-Wallisch C, Dunkler D, for TG2 of the STRATOS initiative. Evaluating variable selection methods for multivariable regression models: A simulation study protocol. PLoS ONE. 2024;19:e0308543. https://doi.org/10.1371/journal.pone.0308543.

20. Palmer SN, Mishra A, Gan S, Liu D, Koh AY, Zhan X. Identifying optimal machine learning approaches for human gut microbiome (shotgun metagenomics) and metabolomics integration with stable feature selection. bioRxiv. 2025. https://doi.org/10.1101/2025.06.21.660858.

21. Takefuji Y. Beyond XGBoost and SHAP: Unveiling true feature importance. Journal of Hazardous Materials. 2025;488:137382. https://doi.org/10.1016/j.jhazmat.2025.137382.

22. Fu L, Zhang H, Dai Y, Zhang H, Pan X, Chen S, et al. Revealing metabolic alterations in brucellosis patients by targeted metabolomics. Journal of Pharmaceutical and Biomedical Analysis. 2024;249:116370. https://doi.org/10.1016/j.jpba.2024.116370.

23. Yu L, Wen M, Ding H. Metabolic map of brucellosis: From acute and chronic phase adaptation to targeted nanodrug design. Materials Today Bio. 2026;36:102684. https://doi.org/10.1016/j.mtbio.2025.102684.

24. Arslan Y, Baran Aİ, Çelik M. Brucellosis-associated hepatitis. Ir J Med Sci. 2024;193:149–56. https://doi.org/10.1007/s11845-023-03382-x.

25. Xu M, Mo S, Fu X. *Brucella* infection-induced hemophagocytic syndrome with subsequent development of the probable vanishing bile duct syndrome: A case report and literature review. SAGE Open Medical Case Reports. 2023;11:2050313X231207562. https://doi.org/10.1177/2050313X231207562.

26. Arriola Benitez PC, Pesce Viglietti AI, Elizalde MM, Giambartolomei GH, Quarleri JF, Delpino MV. Hepatic stellate cells and hepatocytes as liver antigen-presenting cells during B. abortus infection. Pathogens. 2020;9:527. https://doi.org/10.3390/pathogens9070527.

27. Hosseini SM, Farmany A, Arabestani MR. Effect of doxycycline-loaded solid lipid nanoparticles on serum level of trace elements, biochemical and hematological parameters in acute and chronic brucellosis. Biol Trace Elem Res. 2020;194:463–71. https://doi.org/10.1007/s12011-019-01798-0.

28. Manickam K, Kannan S. Brucella spp. — emerging insights into virulence strategies: T4SS, CβG, intracellular replication, and beyond. Journal of Basic Microbiology. 2026;66:e70167. https://doi.org/10.1002/jobm.70167.

29. Elsayed AA, Sallam AM. Nucleotide sequence variants, gene expression and serum profile of immune and antioxidant markers associated with brucellosis resistance/susceptibility in shami goat. Ir Vet J. 2025;78:2. https://doi.org/10.1186/s13620-025-00285-4.

30. Yu J, Yuan H, Guo J, Dong Z, Li S, Fu Q, et al. Combining multi-omics analysis to identify host-targeted targets for the control of *brucella* infection. Microbial Biotechnology. 2023;16:2345–66. https://doi.org/10.1111/1751-7915.14307.

Figures

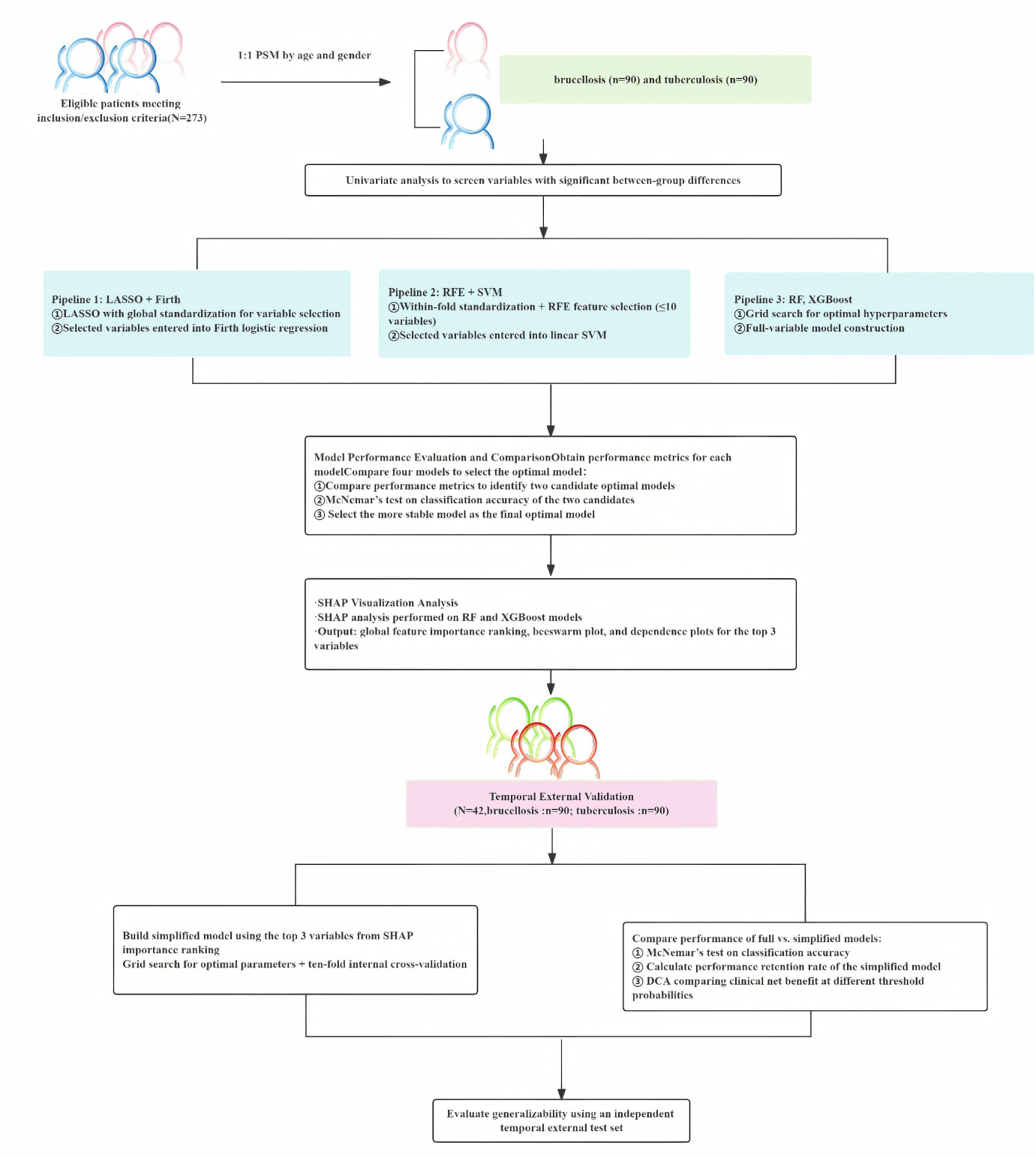


Figure 1

Enrollment process flowchart. Eligible patients (N = 273) were matched 1:1 by propensity score matching (PSM) based on age and sex, resulting in 90 patients with brucellosis and 90 with tuberculosis. Candidate variables were screened via univariate analysis. Four machine learning models (LASSO-Firth

logistic regression, RFE-SVM, Random Forest [RF], and XGBoost) were constructed and compared to select the optimal model. SHAP analysis guided the development of a simplified three-variable model, which was validated using internal cross-validation and an independent temporal external test set.

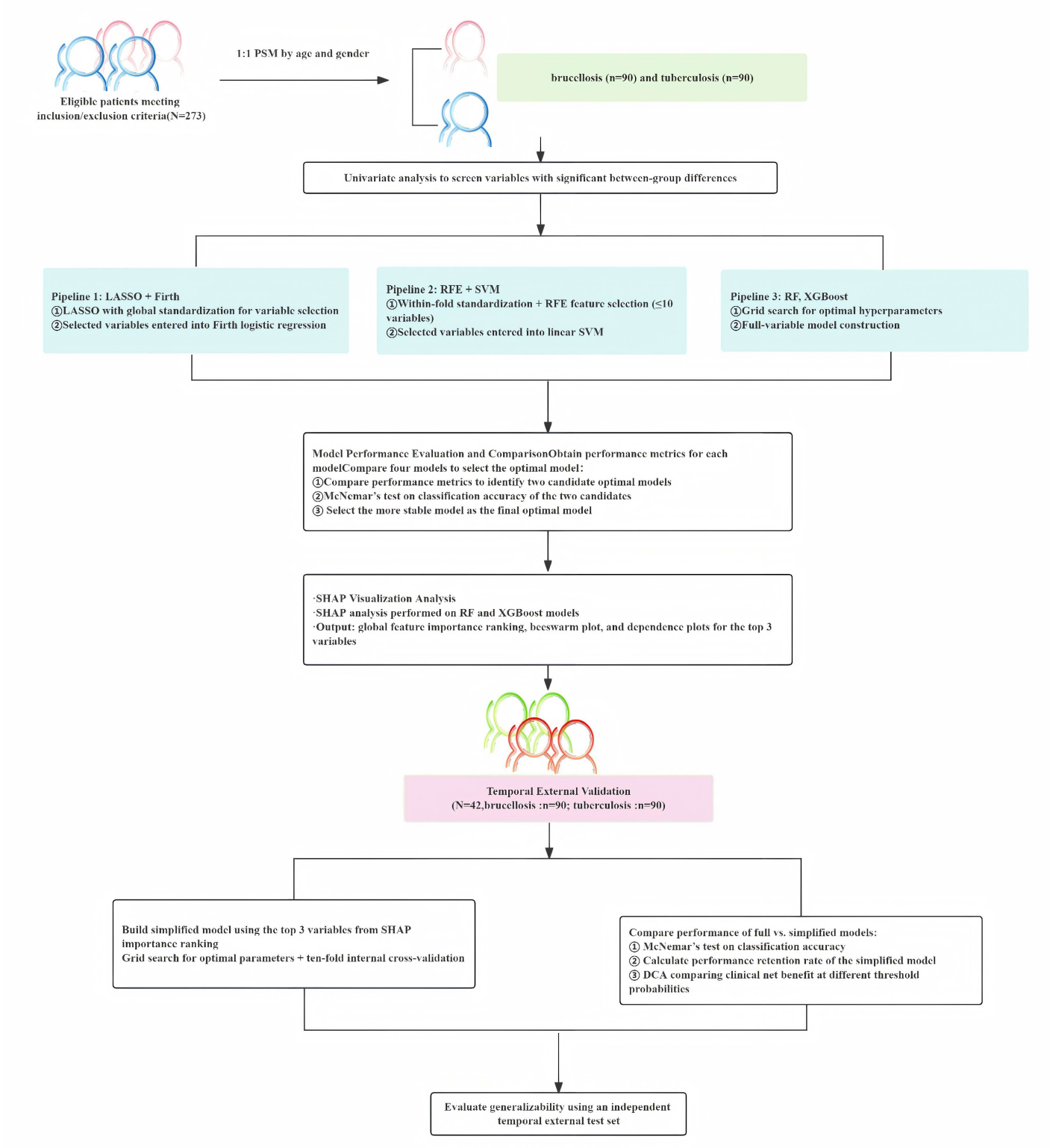


Figure 1

Enrollment process flowchart. Eligible patients (N = 273) were matched 1:1 by propensity score matching (PSM) based on age and sex, resulting in 90 patients with brucellosis and 90 with tuberculosis. Candidate variables were screened via univariate analysis. Four machine learning models (LASSO-Firth logistic regression, RFE-SVM, Random Forest [RF], and XGBoost) were constructed and compared to select the optimal model. SHAP analysis guided the development of a simplified three-variable model, which was validated using internal cross-validation and an independent temporal external test set.

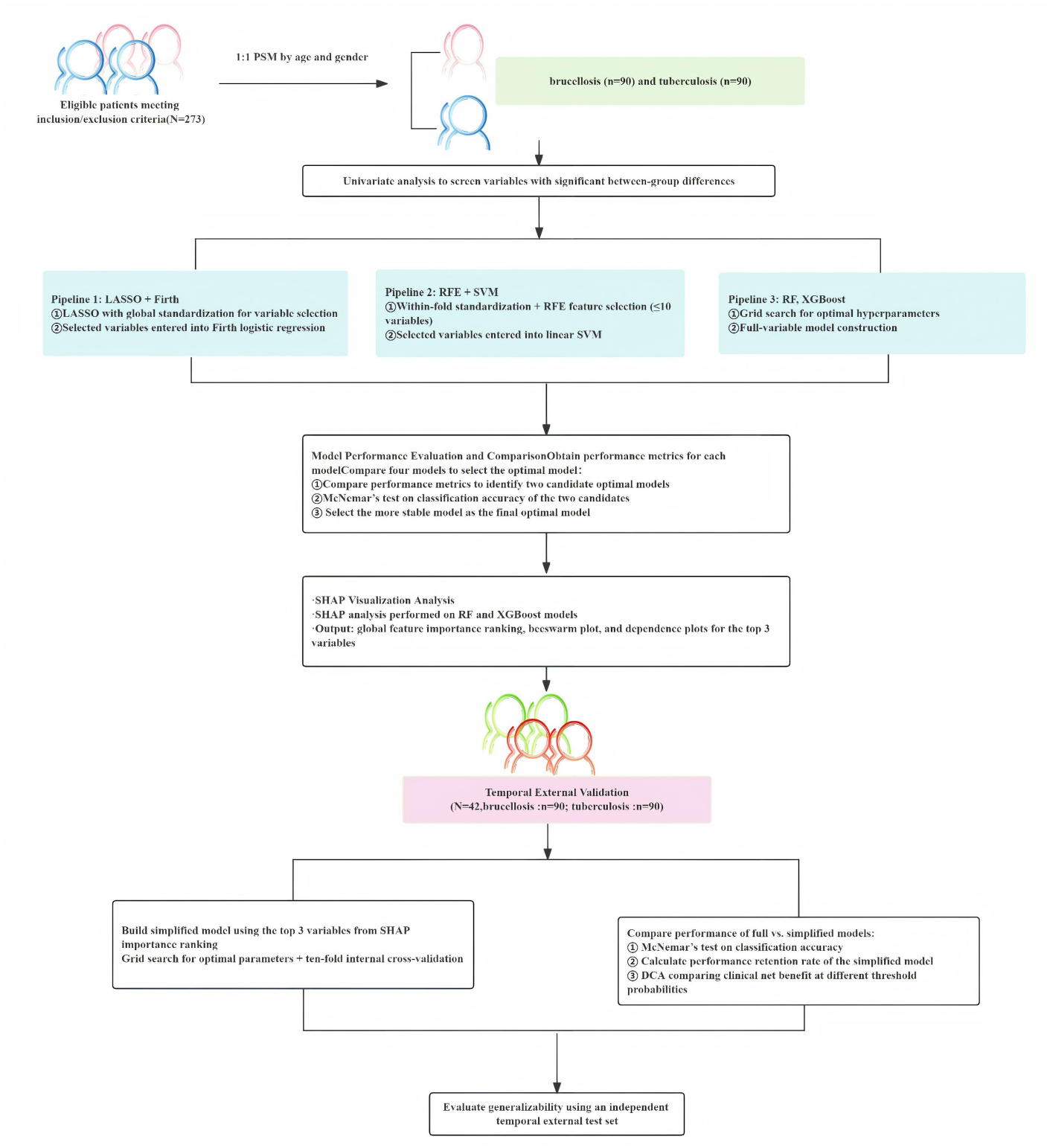


Figure 1

Enrollment process flowchart. Eligible patients (N = 273) were matched 1:1 by propensity score matching (PSM) based on age and sex, resulting in 90 patients with brucellosis and 90 with tuberculosis. Candidate variables were screened via univariate analysis. Four machine learning models (LASSO-Firth logistic regression, RFE-SVM, Random Forest [RF], and XGBoost) were constructed and compared to select the optimal model. SHAP analysis guided the development of a simplified three-variable model, which was validated using internal cross-validation and an independent temporal external test set.

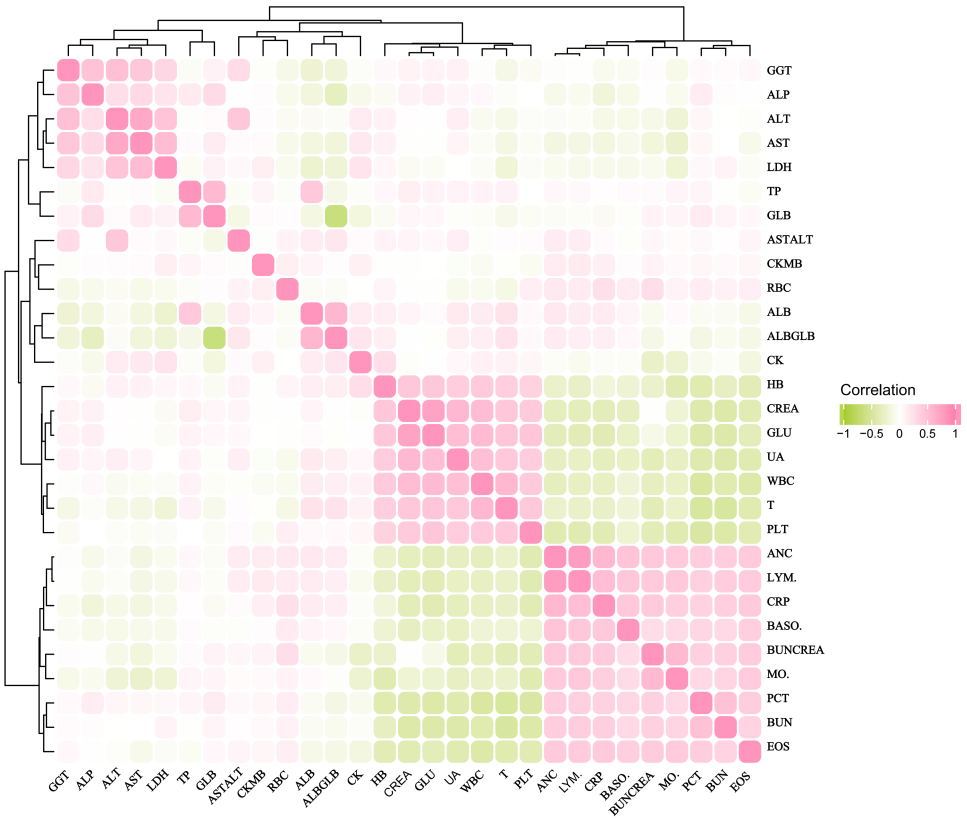


Figure 2

Correlation matrix of routine clinical laboratory biomarkers. Pairwise Pearson correlation coefficients between 29 continuous laboratory parameters are visualized as a heatmap. Each cell represents the correlation coefficient between the corresponding row and column variables. The colour scale denotes the correlation strength: pink (positive correlation) and green (negative correlation), with deeper colours indicating stronger correlations (range: -1 to 1). Hierarchical clustering dendrograms are shown on the top and left to illustrate the similarity of correlation profiles across parameters. Absolute correlation coefficients for all variables were below 0.8, indicating no severe multicollinearity.

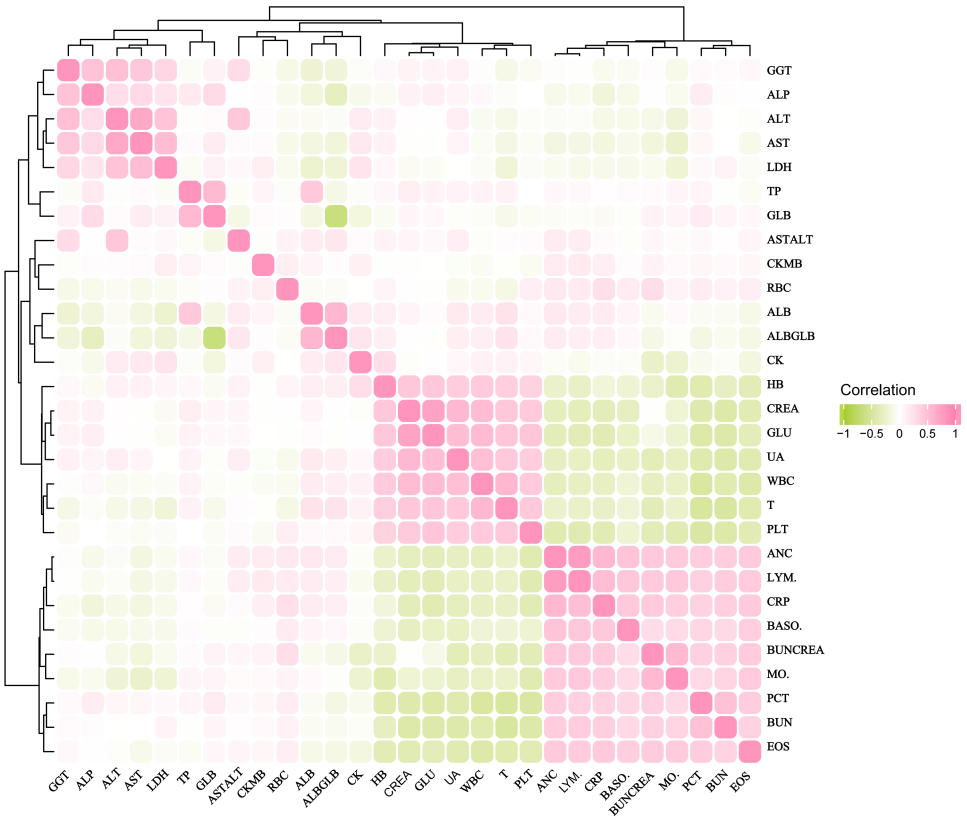


Figure 2

Correlation matrix of routine clinical laboratory biomarkers. Pairwise Pearson correlation coefficients between 29 continuous laboratory parameters are visualized as a heatmap. Each cell represents the correlation coefficient between the corresponding row and column variables. The colour scale denotes the correlation strength: pink (positive correlation) and green (negative correlation), with deeper colours indicating stronger correlations (range: -1 to 1). Hierarchical clustering dendrograms are shown on the top and left to illustrate the similarity of correlation profiles across parameters. Absolute correlation coefficients for all variables were below 0.8, indicating no severe multicollinearity.

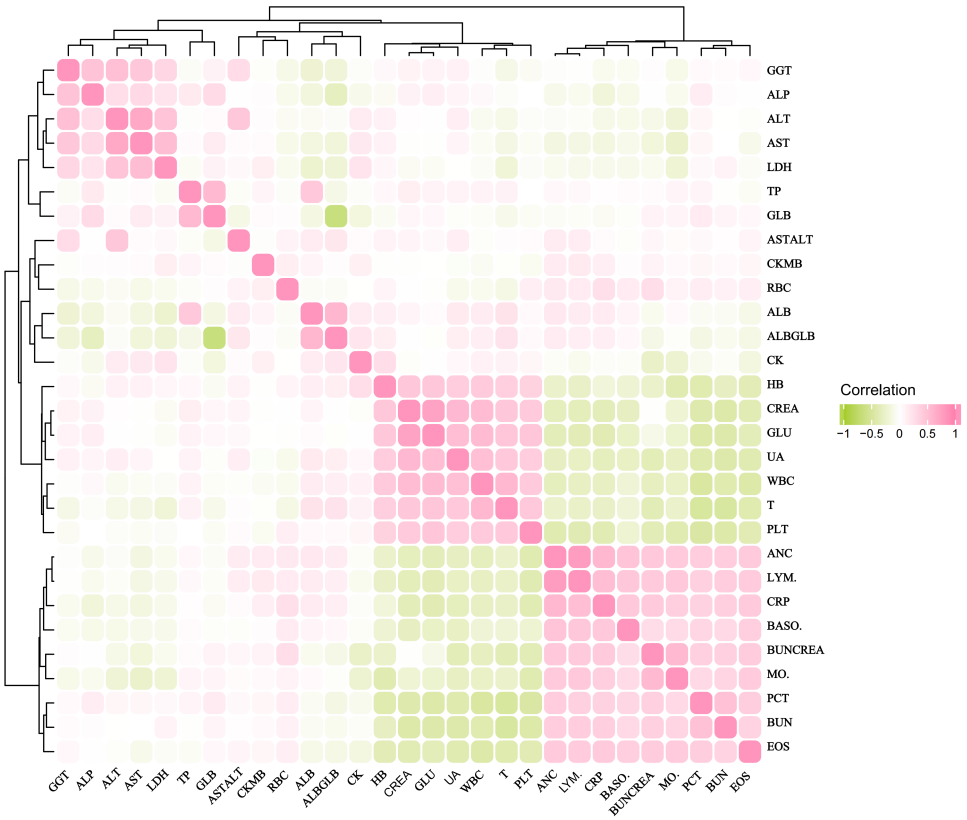


Figure 2

Correlation matrix of routine clinical laboratory biomarkers. Pairwise Pearson correlation coefficients between 29 continuous laboratory parameters are visualized as a heatmap. Each cell represents the correlation coefficient between the corresponding row and column variables. The colour scale denotes the correlation strength: pink (positive correlation) and green (negative correlation), with deeper colours indicating stronger correlations (range: -1 to 1). Hierarchical clustering dendrograms are shown on the top and left to illustrate the similarity of correlation profiles across parameters. Absolute correlation coefficients for all variables were below 0.8, indicating no severe multicollinearity.

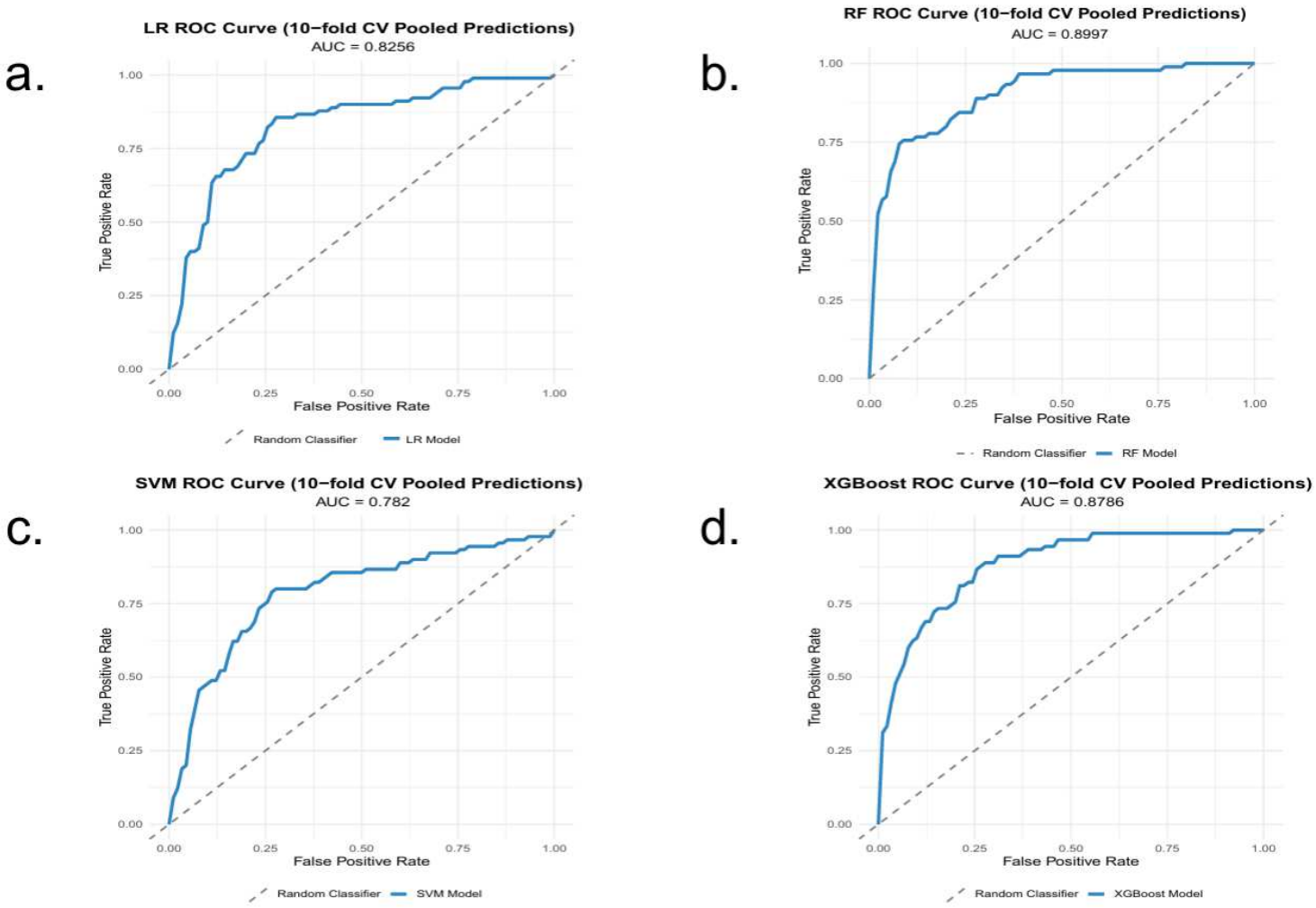


Figure 3

Ten-fold cross-validation ROC curves for the four machine learning models. (a) Logistic regression (LR), (b) Random Forest (RF), (c) support vector machine (SVM), and (d) XGBoost. For each model, the thin grey lines represent the ROC curves of individual folds, and the thick blue line represents the pooled ROC curve. The mean area under the curve (AUC) values were 0.8272 for LR, 0.9056 for RF, 0.7901 for SVM, and 0.8617 for XGBoost. RF achieved the highest AUC and the smallest cross-fold variability, indicating superior discriminative ability and generalisation stability.

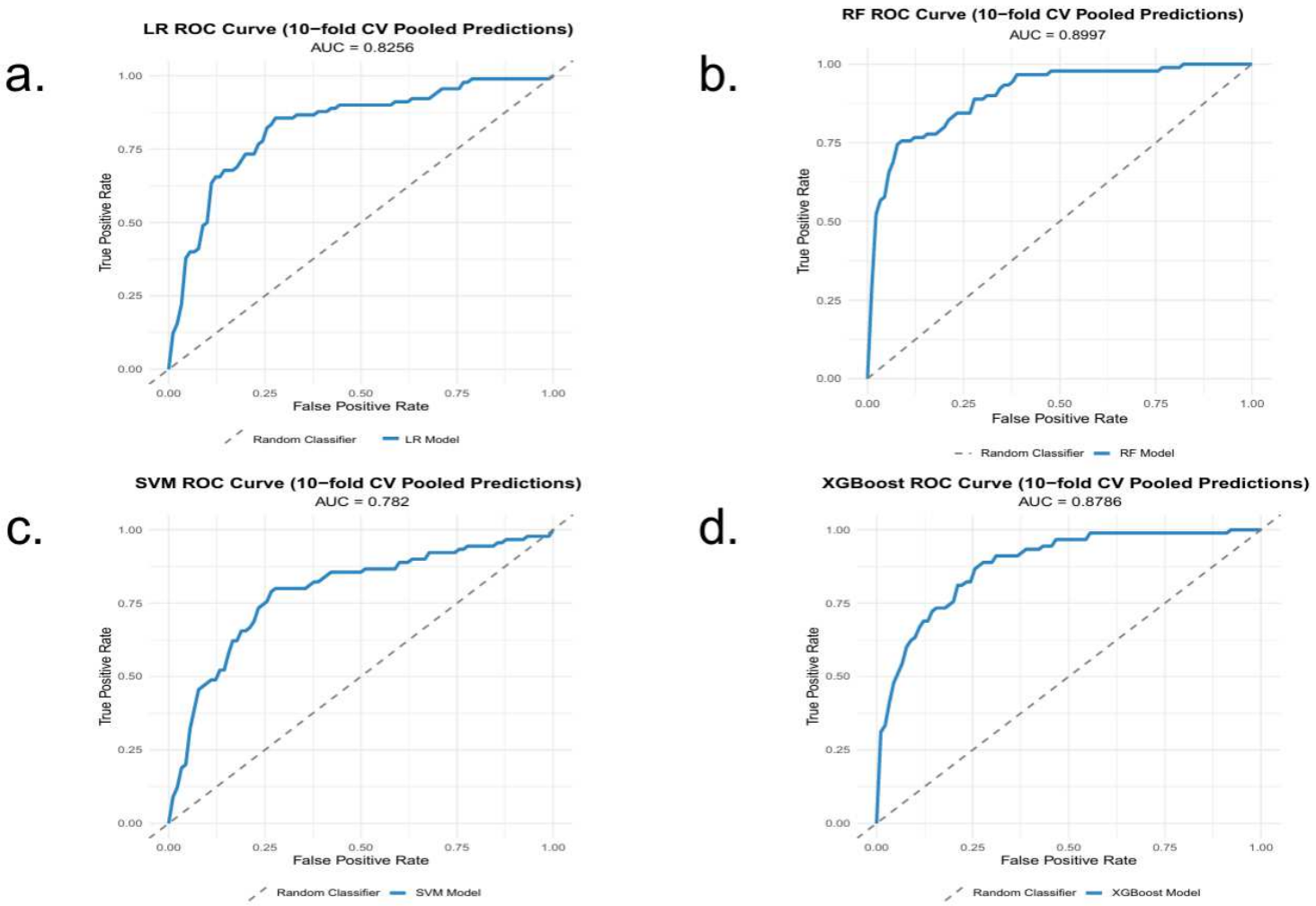


Figure 3

Ten-fold cross-validation ROC curves for the four machine learning models. (a) Logistic regression (LR), (b) Random Forest (RF), (c) support vector machine (SVM), and (d) XGBoost. For each model, the thin grey lines represent the ROC curves of individual folds, and the thick blue line represents the pooled ROC curve. The mean area under the curve (AUC) values were 0.8272 for LR, 0.9056 for RF, 0.7901 for SVM, and 0.8617 for XGBoost. RF achieved the highest AUC and the smallest cross-fold variability, indicating superior discriminative ability and generalisation stability.

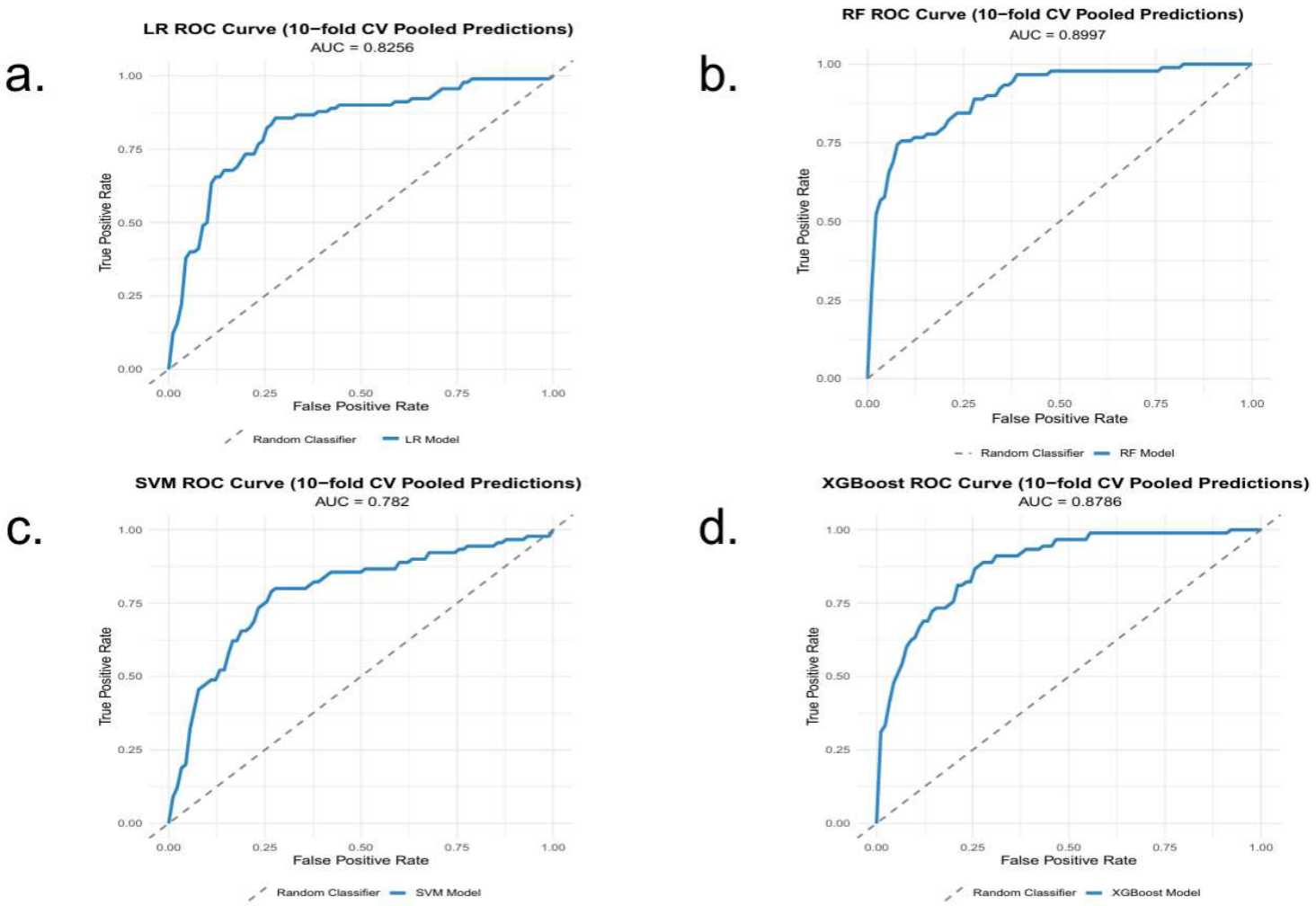


Figure 3

Ten-fold cross-validation ROC curves for the four machine learning models. (a) Logistic regression (LR), (b) Random Forest (RF), (c) support vector machine (SVM), and (d) XGBoost. For each model, the thin grey lines represent the ROC curves of individual folds, and the thick blue line represents the pooled ROC curve. The mean area under the curve (AUC) values were 0.8272 for LR, 0.9056 for RF, 0.7901 for SVM, and 0.8617 for XGBoost. RF achieved the highest AUC and the smallest cross-fold variability, indicating superior discriminative ability and generalisation stability.

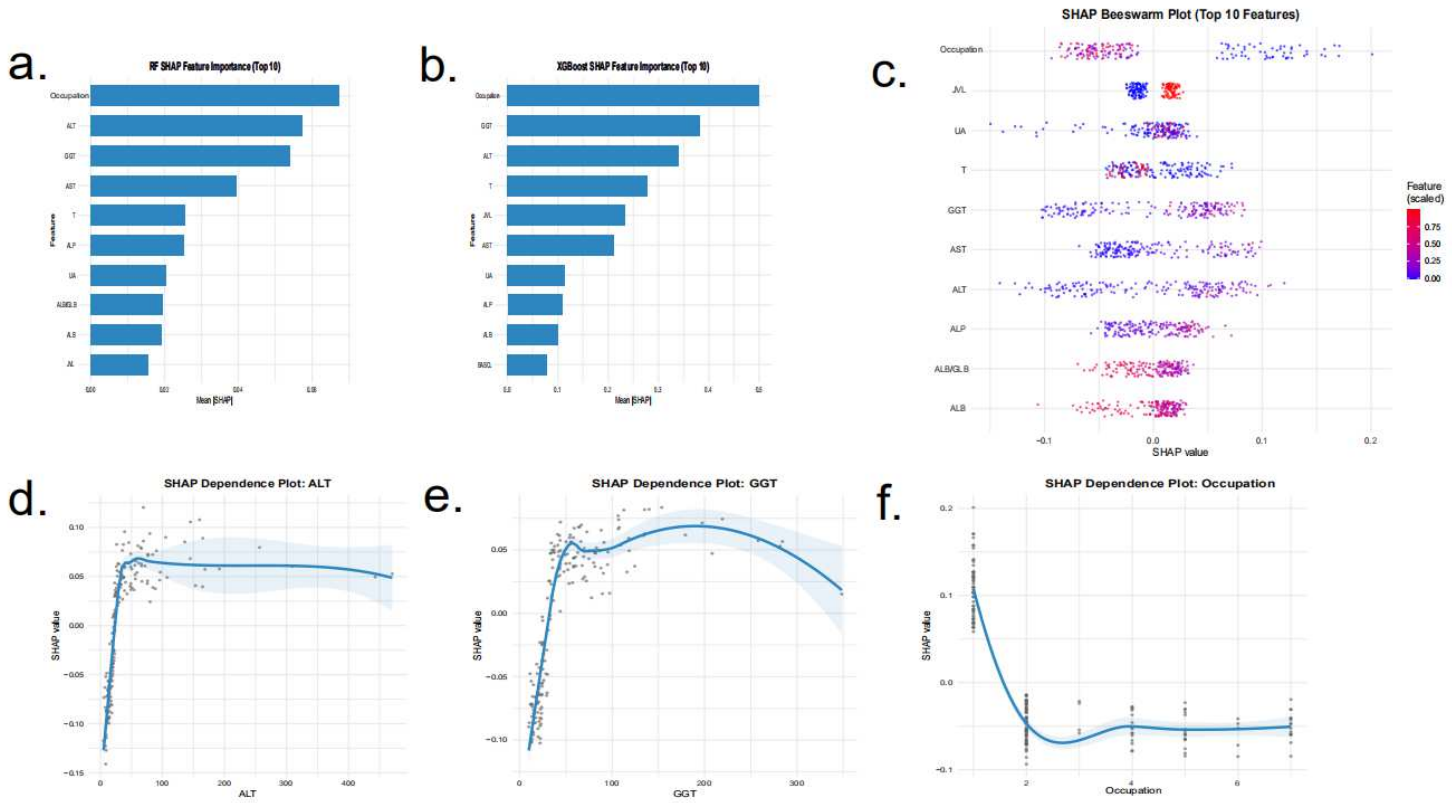


Figure 4

SHAP analysis of the predictive model. (a) The top ten features ranked by mean absolute SHAP values in the Random Forest (RF) model. (b) The top ten features in the XGBoost model. Both models identified occupation, alanine aminotransferase (ALT), and γ-glutamyl transferase (GGT) as the three core discriminative features. (c) SHAP beeswarm plot for the RF model showing the distribution and impact of the top ten features on model output. Each point represents a single patient, coloured by the scaled feature value (blue = low, red = high). Positive SHAP values drive prediction toward brucellosis, and negative values toward tuberculosis. (d–f) SHAP dependence plots for ALT (d), GGT (e), and occupation (f). The blue line represents the trend of the average SHAP value, and the shaded area indicates the confidence interval. Farmer/herder occupation, high ALT, and high GGT positively contributed to brucellosis prediction.

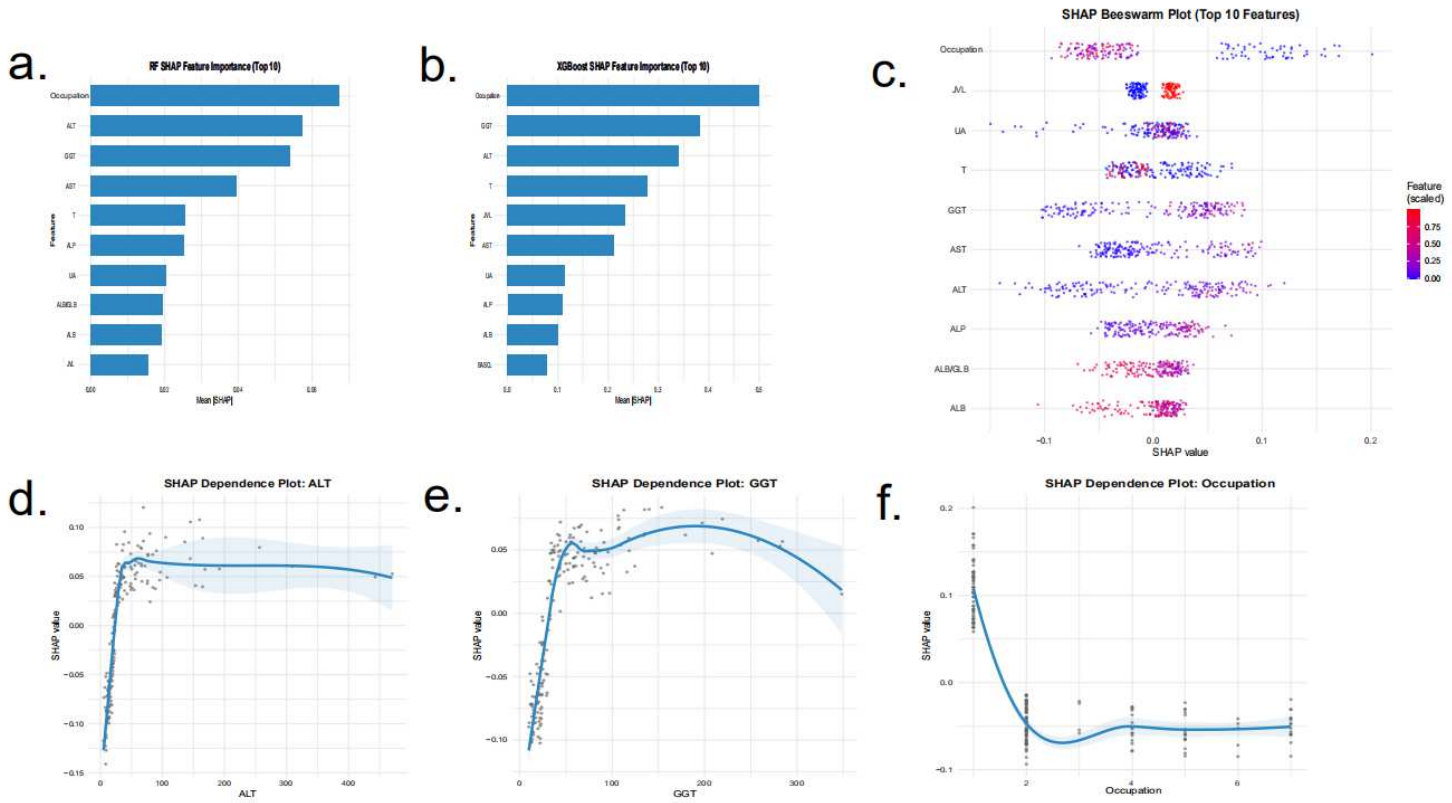


Figure 4

SHAP analysis of the predictive model. (a) The top ten features ranked by mean absolute SHAP values in the Random Forest (RF) model. (b) The top ten features in the XGBoost model. Both models identified occupation, alanine aminotransferase (ALT), and γ-glutamyl transferase (GGT) as the three core discriminative features. (c) SHAP beeswarm plot for the RF model showing the distribution and impact of the top ten features on model output. Each point represents a single patient, coloured by the scaled feature value (blue = low, red = high). Positive SHAP values drive prediction toward brucellosis, and negative values toward tuberculosis. (d–f) SHAP dependence plots for ALT (d), GGT (e), and occupation (f). The blue line represents the trend of the average SHAP value, and the shaded area indicates the confidence interval. Farmer/herder occupation, high ALT, and high GGT positively contributed to brucellosis prediction.

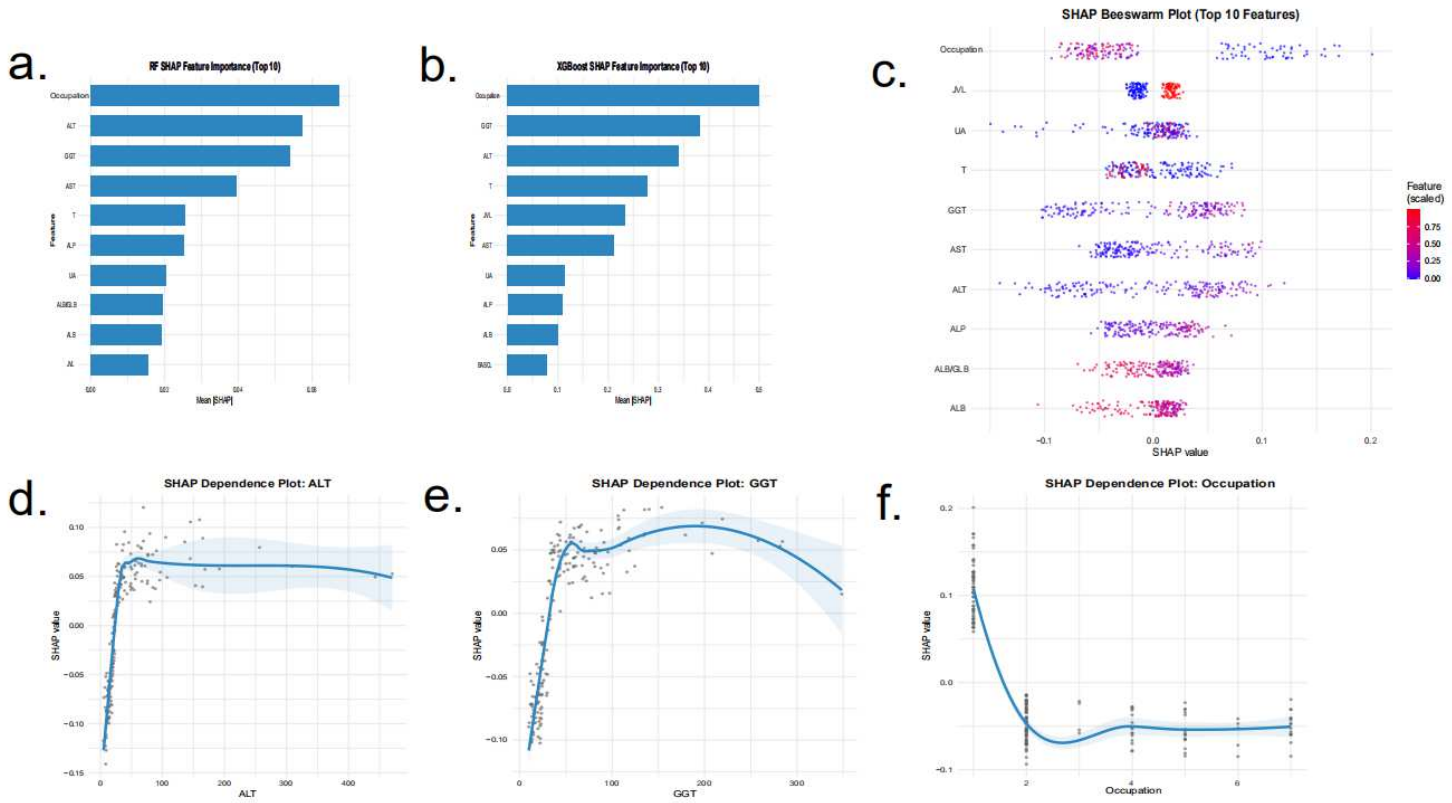


Figure 4

SHAP analysis of the predictive model. (a) The top ten features ranked by mean absolute SHAP values in the Random Forest (RF) model. (b) The top ten features in the XGBoost model. Both models identified occupation, alanine aminotransferase (ALT), and γ-glutamyl transferase (GGT) as the three core discriminative features. (c) SHAP beeswarm plot for the RF model showing the distribution and impact of the top ten features on model output. Each point represents a single patient, coloured by the scaled feature value (blue = low, red = high). Positive SHAP values drive prediction toward brucellosis, and negative values toward tuberculosis. (d–f) SHAP dependence plots for ALT (d), GGT (e), and occupation (f). The blue line represents the trend of the average SHAP value, and the shaded area indicates the confidence interval. Farmer/herder occupation, high ALT, and high GGT positively contributed to brucellosis prediction.

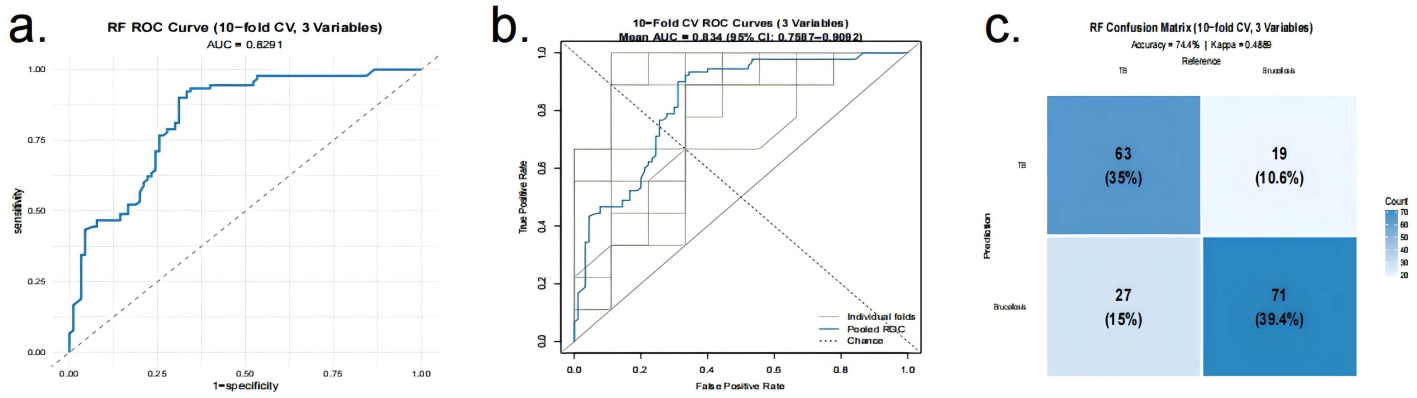


Figure 5

Performance evaluation of the simplified three-variable Random Forest (RF) model. (a) Receiver operating characteristic (ROC) curve of the simplified RF model on the full dataset (N = 180), with an area under the curve (AUC) of 0.8291. (b) Ten-fold cross-validation ROC curves. The blue line represents the pooled ROC curve (mean AUC = 0.834, 95% CI: 0.7587–0.9092), and the grey lines denote the ROC curves of individual folds. (c) Confusion matrix of the simplified RF model. The overall accuracy was 74.4%, and the Kappa coefficient was 0.4889. TB, tuberculosis.

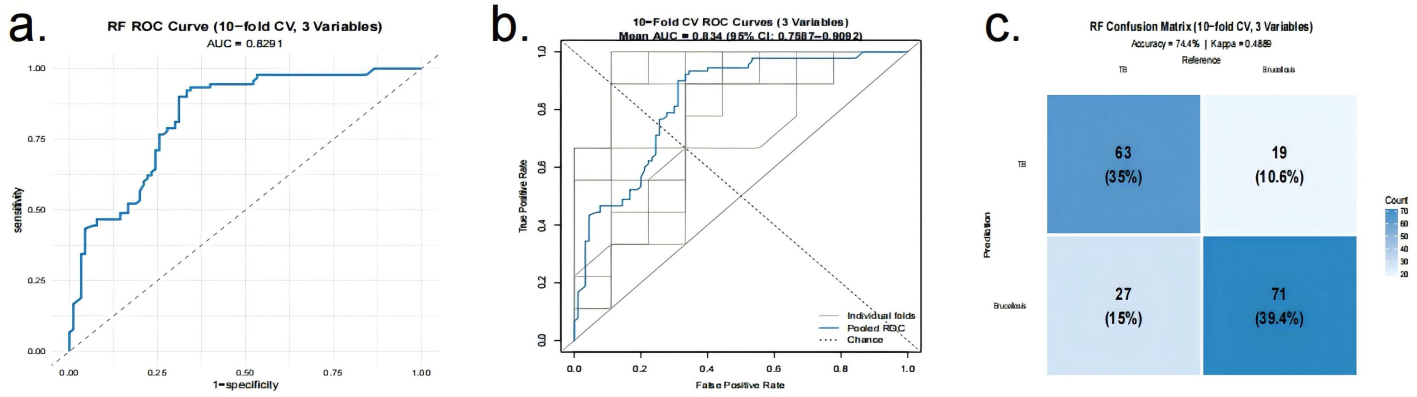


Figure 5

Performance evaluation of the simplified three-variable Random Forest (RF) model. (a) Receiver operating characteristic (ROC) curve of the simplified RF model on the full dataset (N = 180), with an area under the curve (AUC) of 0.8291. (b) Ten-fold cross-validation ROC curves. The blue line represents the pooled ROC curve (mean AUC = 0.834, 95% CI: 0.7587–0.9092), and the grey lines denote the ROC

curves of individual folds. (c) Confusion matrix of the simplified RF model. The overall accuracy was 74.4%, and the Kappa coefficient was 0.4889. TB, tuberculosis.

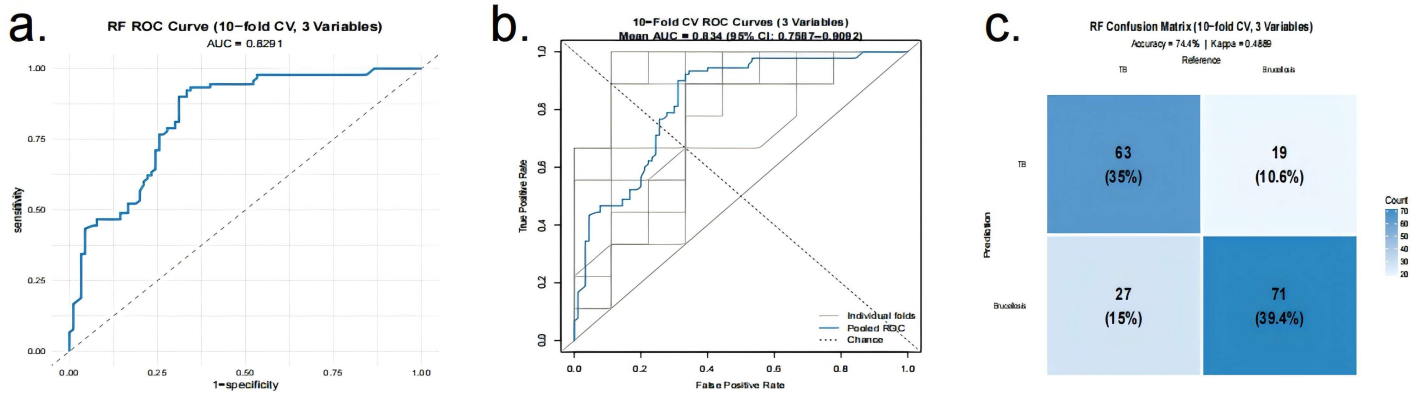


Figure 5

Performance evaluation of the simplified three-variable Random Forest (RF) model. (a) Receiver operating characteristic (ROC) curve of the simplified RF model on the full dataset (N = 180), with an area under the curve (AUC) of 0.8291. (b) Ten-fold cross-validation ROC curves. The blue line represents the pooled ROC curve (mean AUC = 0.834, 95% CI: 0.7587–0.9092), and the grey lines denote the ROC curves of individual folds. (c) Confusion matrix of the simplified RF model. The overall accuracy was 74.4%, and the Kappa coefficient was 0.4889. TB, tuberculosis.

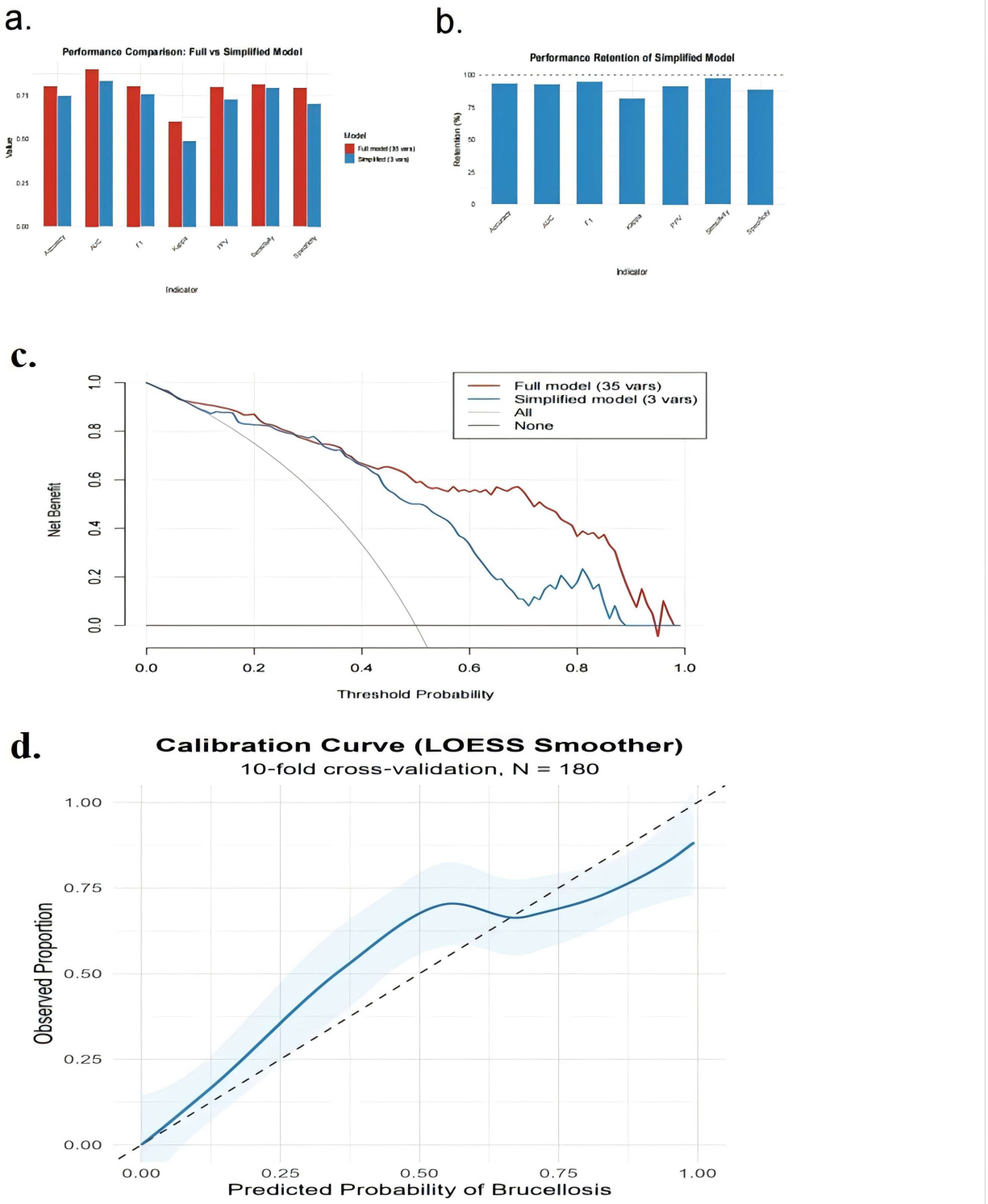


Figure 6

Performance comparison and clinical utility of the full (35-variable) and simplified (3-variable) models. (a) Comparison of key performance metrics (accuracy, AUC, F1 score, Kappa, positive predictive value [PPV], sensitivity, and specificity) between the full and simplified models. (b) Performance retention of the simplified model relative to the full model, expressed as the percentage of each metric retained. Sensitivity, AUC, accuracy, and F1 score retention exceeded 90%. (c) Decision curve analysis (DCA)

showing the net benefit of the full model (red line) and the simplified model (blue line) across different threshold probabilities. The grey lines represent the "treat all" and "treat none" strategies. The two curves largely overlap within the clinically reasonable threshold range (0–0.4). (d) Calibration curve (LOESS smoother) of the simplified model for brucellosis prediction, based on ten-fold cross-validation (N = 180). The blue line and shaded area represent the observed proportion and its 95% confidence interval, while the dashed line indicates perfect calibration. The fitted curve aligns well with the ideal calibration line.

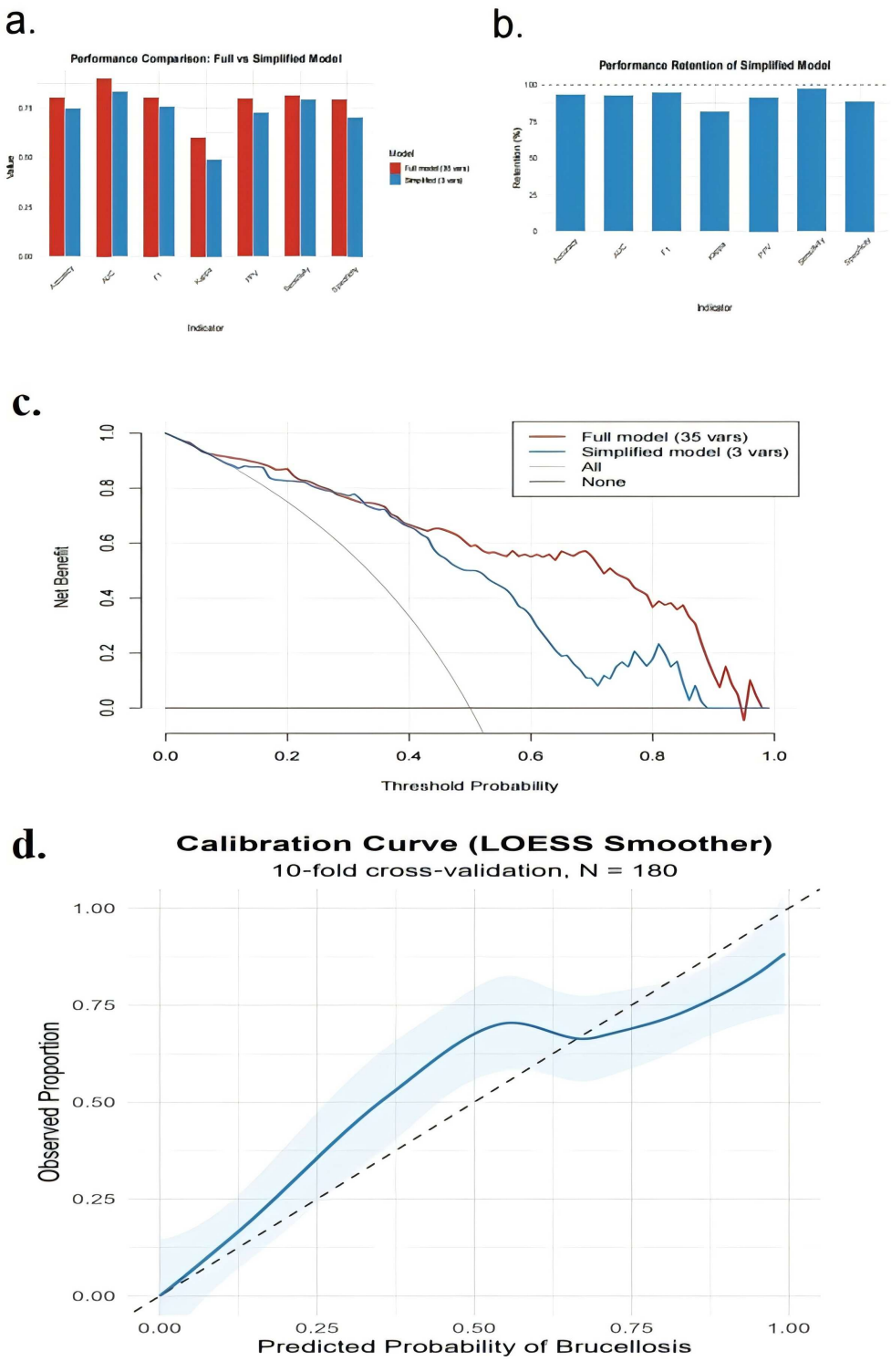


Figure 6

Performance comparison and clinical utility of the full (35-variable) and simplified (3-variable) models. (a) Comparison of key performance metrics (accuracy, AUC, F1 score, Kappa, positive predictive value [PPV], sensitivity, and specificity) between the full and simplified models. (b) Performance retention of the simplified model relative to the full model, expressed as the percentage of each metric retained. Sensitivity, AUC, accuracy, and F1 score retention exceeded 90%. (c) Decision curve analysis (DCA) showing the net benefit of the full model (red line) and the simplified model (blue line) across different threshold probabilities. The grey lines represent the "treat all" and "treat none" strategies. The two curves largely overlap within the clinically reasonable threshold range (0–0.4). (d) Calibration curve (LOESS smoother) of the simplified model for brucellosis prediction, based on ten-fold cross-validation (N = 180). The blue line and shaded area represent the observed proportion and its 95% confidence interval, while the dashed line indicates perfect calibration. The fitted curve aligns well with the ideal calibration line.

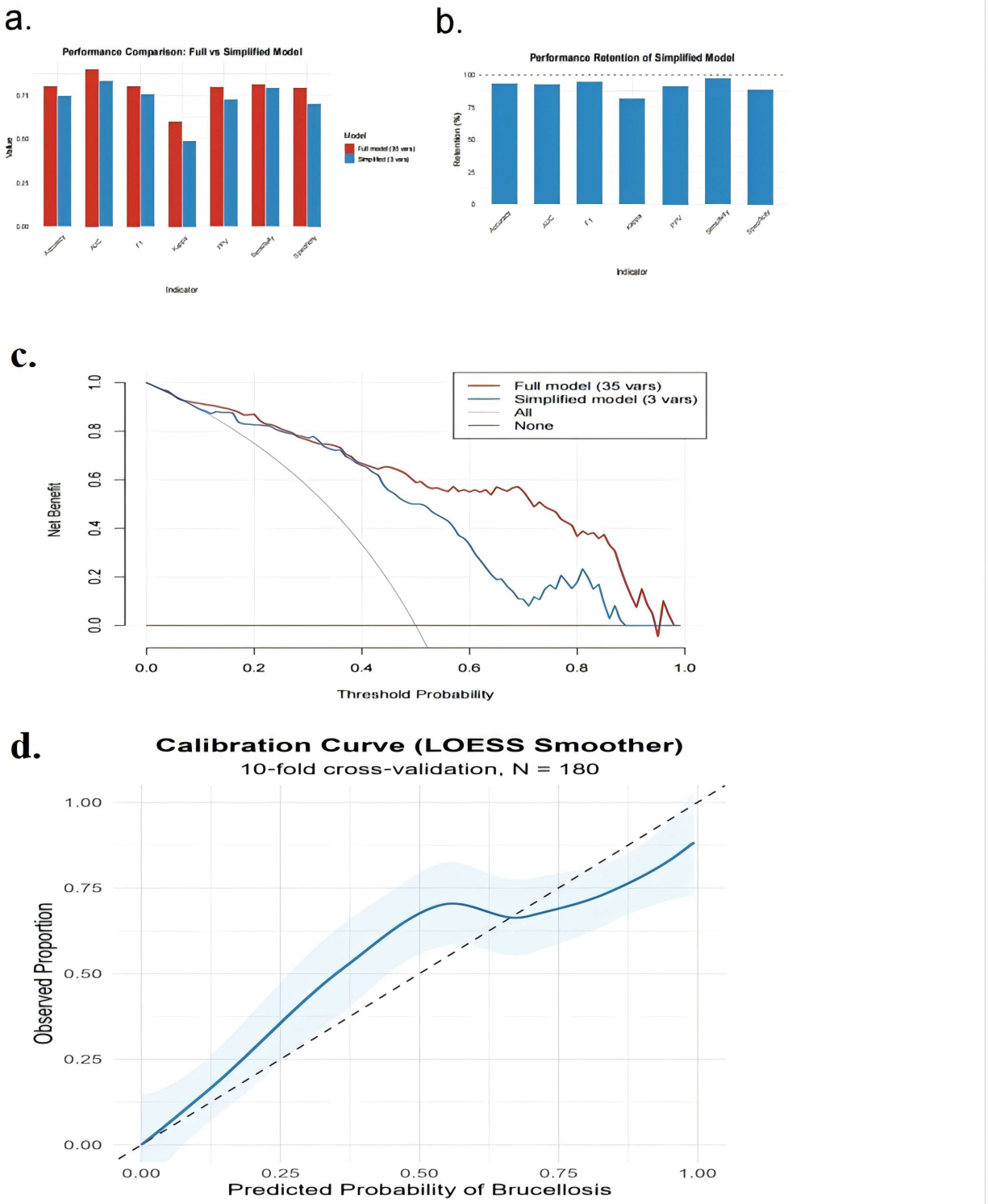


Figure 6

Performance comparison and clinical utility of the full (35-variable) and simplified (3-variable) models. (a) Comparison of key performance metrics (accuracy, AUC, F1 score, Kappa, positive predictive value [PPV], sensitivity, and specificity) between the full and simplified models. (b) Performance retention of the simplified model relative to the full model, expressed as the percentage of each metric retained. Sensitivity, AUC, accuracy, and F1 score retention exceeded 90%. (c) Decision curve analysis (DCA)

showing the net benefit of the full model (red line) and the simplified model (blue line) across different threshold probabilities. The grey lines represent the "treat all" and "treat none" strategies. The two curves largely overlap within the clinically reasonable threshold range (0–0.4). (d) Calibration curve (LOESS smoother) of the simplified model for brucellosis prediction, based on ten-fold cross-validation (N = 180). The blue line and shaded area represent the observed proportion and its 95% confidence interval, while the dashed line indicates perfect calibration. The fitted curve aligns well with the ideal calibration line.

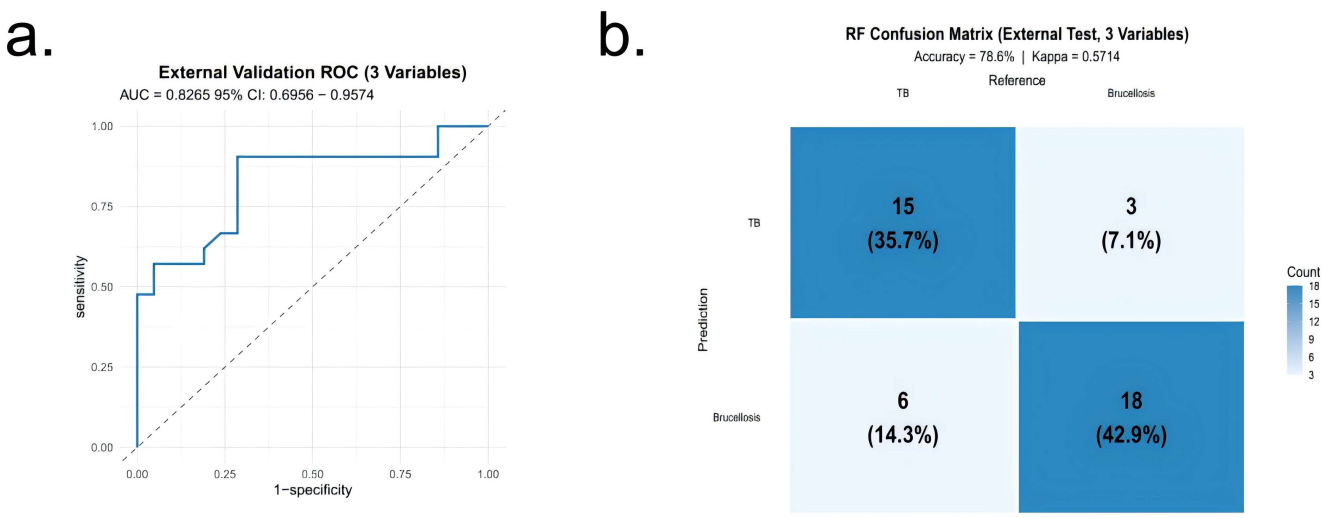


Figure 7

External validation of the simplified three-variable Random Forest (RF) model on a temporally independent test set (n = 42). (a) Receiver operating characteristic (ROC) curve. The area under the curve (AUC) was 0.8265 (95% CI: 0.6956–0.9574). (b) Confusion matrix. The overall accuracy was 78.6%, sensitivity 85.7%, specificity 71.4%, and Kappa coefficient 0.5714, indicating moderate agreement between model predictions and true labels. TB, tuberculosis.

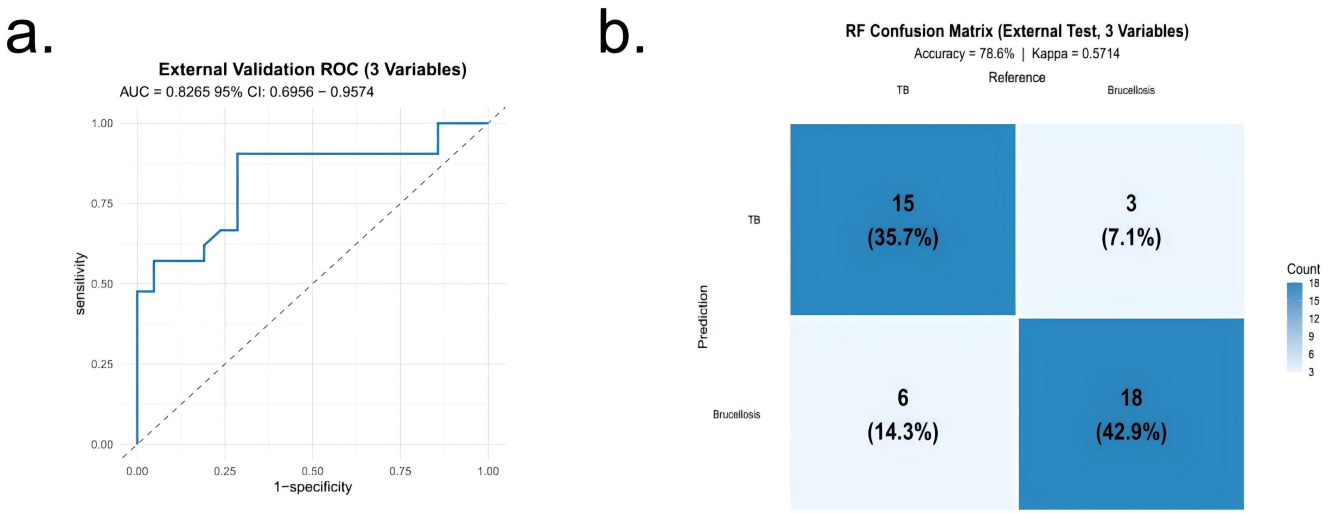


Figure 7

External validation of the simplified three-variable Random Forest (RF) model on a temporally independent test set (n = 42). (a) Receiver operating characteristic (ROC) curve. The area under the curve (AUC) was 0.8265 (95% CI: 0.6956–0.9574). (b) Confusion matrix. The overall accuracy was 78.6%, sensitivity 85.7%, specificity 71.4%, and Kappa coefficient 0.5714, indicating moderate agreement between model predictions and true labels. TB, tuberculosis.

a.
External Validation ROC (3 Variables)
AUC = 0.8265 95% CI: 0.6956 – 0.9574
sensitivity
1−specificity
0.00
0.25
0.50
0.75
1.00
b.
RF Confusion Matrix (External Test, 3 Variables)
Accuracy = 78.6% | Kappa = 0.5714
Reference
TB
Brucellosis
Prediction
15
(35.7%)
3
(7.1%)
6
(14.3%)
18
(42.9%)
Count
18
15
12
9
6
3

Figure 7

External validation of the simplified three-variable Random Forest (RF) model on a temporally independent test set (n = 42). (a) Receiver operating characteristic (ROC) curve. The area under the curve (AUC) was 0.8265 (95% CI: 0.6956–0.9574). (b) Confusion matrix. The overall accuracy was 78.6%, sensitivity 85.7%, specificity 71.4%, and Kappa coefficient 0.5714, indicating moderate agreement between model predictions and true labels. TB, tuberculosis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.zip
- table1.xlsx
- SupplementaryMaterials.zip
- table1.xlsx
- table1.xlsx
- SupplementaryMaterials.zip
- table2.xlsx
- table2.xlsx
- table2.xlsx
- table3.xlsx
- table3.xlsx
- table3.xlsx
- table4.xlsx
- table4.xlsx
- table4.xlsx
- table5.xlsx
- table5.xlsx
- table5.xlsx