

Development of a Machine-Learning Immuno-Serologic Diagnostic Model for Non-Neutropenic Invasive Pulmonary Fungal Disease

Hui Huang^{1,2,*}, Fang Fang^{3,*}, Weiguo Lu², Zhihui Liu^{2,4}, Junyuan Huang² 

¹The First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, People's Republic of China; ²Department of Clinical Laboratory Medicine, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, People's Republic of China; ³Department of Clinical Laboratory Medicine, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, People's Republic of China; ⁴Shenshan Hospital, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Shanwei, Guangdong, People's Republic of China

*These authors contributed equally to this work

Correspondence: Zhihui Liu; Junyuan Huang, Email liuzhihui1593@gzucm.edu.cn; juyua1989@163.com

Background: Non-neutropenic invasive pulmonary fungal disease (IPFD) is increasingly recognized but remains challenging to diagnose due to nonspecific clinical and radiological features. This retrospective, single-center study was conducted at the First Affiliated Hospital of Guangzhou University of Chinese Medicine and aimed to develop and evaluate a diagnostic model based on immuno-Serologic biomarkers for distinguishing non-neutropenic IPFD from bacterial pneumonia.

Methods: A total of 157 pneumonia patients (65 non-neutropenic IPFD cases and 92 bacterial pneumonia cases) admitted to the First Affiliated Hospital of Guangzhou University of Chinese Medicine between April 2018 and December 2022 were enrolled. Least Absolute Shrinkage and Selection Operator (LASSO) regression and collinearity analysis were applied to screen key variables, followed by the development of diagnostic models using nine machine learning algorithms. Model performance was comprehensively evaluated, and temporal validation in an independent later cohort from the same center was conducted using data from 102 pneumonia patients (33 non-neutropenic IPFD and 69 bacterial pneumonia cases) admitted between January 2023 and March 2025.

Results: Five biomarkers were identified as predictors: galactomannan (GM), monocyte human leukocyte antigen-DR expression (mHLA-DR), monocyte count, interleukin-6 (IL-6), and 1,3- β -D-glucan (BDG). The Light Gradient Boosting Machine (LightGBM) model demonstrated optimal performance in the validation set, with an area under the receiver operating characteristic curve (AUC) of 0.865 (95% CI: 0.728–0.999) and accuracy of 0.781. In the test set, the model achieved an AUC of 0.810 and accuracy of 0.750. Decision curve analysis (DCA) indicated favorable net benefits across probability thresholds of 0–1. Temporal validation yielded an AUC of 0.821 and accuracy of 0.794.

Conclusion: The immuno-serologic diagnostic model exhibits strong discriminatory performance in differentiating bacterial pneumonia from non-neutropenic IPFD, highlighting its potential for clinical application.

Keywords: non-neutropenic invasive pulmonary fungal disease, serological biomarkers, diagnostic model, machine learning

Introduction

In recent years, the incidence of IPFD has been rising annually, posing serious threats to patient health. While it predominantly occurs in neutropenic populations, its incidence has also increased among non-neutropenic individuals.^{1,2} The most common fungal pathogens responsible for IPFD include *Aspergillus* species, *Candida* species, and less frequently, *Mucorales*, *Cryptococcus*, and *Fusarium*.^{1,3} Non-neutropenic IPFD presents complex immune status, atypical clinical manifestations, and nonspecific imaging findings, often leading to misdiagnosis as bacterial pneumonia.

Critically ill patients with delayed treatment frequently experience poor prognosis, making early identification of this population a key challenge in clinical management.⁴ Although predictive models have been applied for clinical prediction and prognosis assessment of invasive fungal infections, such as early diagnostic models for invasive pulmonary aspergillosis based on chest computed tomography (CT) images and risk prediction models using the Medical Information Mart for Intensive Care IV (MIMIC-IV) database,^{5–7} few studies have focused on diagnostic models for non-neutropenic IPFD, particularly those based on immune cell detection.

The “gold standard” for diagnosing non-neutropenic IPFD requires histopathological confirmation of hyphae presence or positive cultures. However, invasive procedures are often contraindicated in many patients due to high bleeding risks or compromised respiratory status, preventing acquisition of biopsy specimens.⁸ Furthermore, bronchoalveolar lavage fluid fungal cultures demonstrate sensitivity of only 30%–60%, coupled with issues including prolonged culture periods and susceptibility to contamination.⁹ Fungal antigen detection methods (eg, BDG, GM, and cryptococcal capsular polysaccharide antigen) show limited sensitivity for diagnosing non-neutropenic IPFD and are susceptible to interference from bacterial infections and mucosal colonization.¹⁰ There is therefore a need for a diagnostic model that provides precise decision support to reduce diagnostic delays in non-neutropenic IPFD and to quantify immune–infection interactions.

This study aims to construct and validate a machine learning-based diagnostic model utilizing readily available serological and immune cell parameters to facilitate the early and accurate distinction of non-neutropenic IPFD from bacterial pneumonia.

Material and Methods

Cohort Selection and Data Collection

This retrospective cohort study consecutively enrolled 157 pneumonia patients admitted to the First Affiliated Hospital of Guangzhou University of Chinese Medicine between April 2018 and December 2022. Participants were divided into a study group (65 non-neutropenic IPFD cases) and a control group (92 bacterial pneumonia cases). An additional temporal validation set comprising an independent later cohort from the same center set of 102 pneumonia patients (33 non-neutropenic IPFD and 69 bacterial pneumonia cases) admitted from January 2023 to March 2025.

All enrolled patients had complete clinical documentation and met the inclusion criteria defined by the Expert Consensus on Diagnosis and Treatment of Pulmonary Fungal Diseases.¹¹ The host risk factors for IPFD considered in this study included: chronic obstructive pulmonary disease (COPD), diabetes mellitus, prolonged antibiotic use, recent major surgery, renal insufficiency, and prolonged intensive care unit (ICU) stay. IPFD cases were categorized into proven IPFD, clinically diagnosed IPFD, and suspected IPFD based on clinical manifestations, microbiological evidence, and pathological findings. This study included proven IPFD and clinically diagnosed IPFD cases in the IPFD group, defined as follows:

(1) Proven IPFD required ≥ 1 host risk factor, clinical features of invasive fungal disease, and histopathological confirmation of fungal tissue invasion and/or positive fungal culture from sterile specimens (eg, lung tissue obtained via sterile procedure, pleural fluid, or blood);

(2) Clinically diagnosed IPFD required ≥ 1 host risk factor, one major or two minor clinical features of invasive fungal disease, direct microscopic identification of fungal hyphae in endotracheal aspirates or qualified sputum samples with ≥ 2 consecutive cultures isolating the same fungal species, or detection of hyphae in bronchoalveolar lavage fluid (BALF) via direct microscopy combined with positive fungal culture.

Among the 65 non-neutropenic IPFD patients, 10 were classified as proven IPFD and 55 as clinically diagnosed IPFD. The control group consisted of 92 patients with bacterial pneumonia, confirmed by positive bacterial culture or molecular testing. In the temporal validation set, among the 33 non-neutropenic IPFD patients, 5 were classified as proven IPFD and 28 as clinically diagnosed IPFD. The control group in the temporal validation set consisted of 69 patients with bacterial pneumonia, confirmed by positive bacterial culture or molecular testing.

Exclusion criteria included: (a) neutropenia (absolute neutrophil count $<0.5 \times 10^9/L$); (b) long-term corticosteroid (equivalent to prednisone ≥ 20 mg/day for ≥ 2 weeks) or immunosuppressive therapy; (c) malignancies or autoimmune diseases; (d) severe concurrent viral infections; and (e) prior antibiotic treatment within 72 hours before admission.

Data Collection

Demographic characteristics (including sex and age), laboratory test results, and supplementary clinical data were retrospectively acquired from the participants' medical records. The immunocyte subset analysis, Cluster of Differentiation 64 (CD64) index, and mHLA-DR detection were performed using flow cytometry (DxFLEX flow cytometer, Beckman Coulter, Inc, USA). BDG was measured through a chromogenic assay (Dynamiker Fungus (1–3)- β -D-Glucan Assay, Dynamiker Biotechnology (Tianjin) Co., Ltd., China), while GM was analyzed via enzyme-linked immunosorbent assay (ELISA) (Dynamiker Aspergillus Galactomannan Assay, Dynamiker Biotechnology (Tianjin) Co., Ltd., China). IL-6 levels were quantified using electrochemiluminescence immunoassay (ECLIA) (Cobas e601, Roche Diagnostics, Germany), and C-reactive protein (CRP) concentrations were determined by nephelometry (Immage 800, Beckman Coulter, Inc, USA).

Feature Selection

Twelve laboratory parameters were analyzed, categorized into fungal antigens (BDG, GM), immune cell subsets (CD3+ T cells, CD4+ T cells, CD8+ T cells, NK cells, B cells, monocytes), inflammatory markers (IL-6, CRP, neutrophil CD64 index), and mHLA-DR. Variable selection was conducted through LASSO regression,¹² with the optimal regularization parameter λ determined via 10-fold cross-validation. The L1-penalty mechanism effectively compressed coefficients of non-significant variables to zero. Subsequent collinearity analysis was performed to assess multicollinearity among the variables selected by LASSO. Multicollinearity, which occurs when predictor variables are highly correlated, can inflate the variance of coefficient estimates and destabilize the model. This was quantified using the Variance Inflation Factor (VIF). Variables with a VIF > 10 , indicating severe multicollinearity, were excluded to further refine the model.¹³ This dual-stage optimization ultimately identified five key predictors: BDG, GM, IL-6, mHLA-DR, and Monocytes count.

Model Development and Validation

Nine machine learning algorithms: Logistic Regression (Logistic), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Random Forest (RF), Adaptive Boosting (AdaBoost), Decision Tree (DT), Gradient Boosting Decision Tree (GBDT), Gaussian Naive Bayes (GNB), and Complement Naive Bayes (CNB) were employed to construct predictive models. Model performance was compared through receiver operating characteristic (ROC) curves and decision curve analysis (DCA) to identify the model with optimal diagnostic performance.^{14,15} Shapley Additive exPlanations (SHAP) plots were used to elucidate variable importance.¹⁶ Temporal validation was performed using the independent validation cohort,¹⁷ ultimately generating an online prediction tool. The overall study design is illustrated in [Figure 1](#).

Statistical Analysis

Statistical analyses were conducted utilizing SPSS 24.0 (IBM Corp) and the Beckman Coulter DxAI 800 analytical system. The normality of continuous variables was assessed using the Shapiro–Wilk test, supplemented by visual inspection of Q-Q plots and histograms. Continuous variables demonstrating normal distribution were analyzed using Student's *t*-test with least significant difference (LSD) post-hoc analysis. For non-normally distributed continuous variables, the Mann–Whitney *U*-test was applied. Categorical variables were assessed through χ^2 -tests or Fisher's exact probability tests as appropriate. Statistical significance was defined as a two-tailed *P*-value < 0.05 across all analyses.

Results

Baseline Characteristics of the Study Participants

Comparative analysis revealed significant between-group differences in laboratory parameters, including BDG, GM, IL-6, CRP, CD64 index, CD4+ T cells, and monocytes ($P < 0.05$). In contrast, no statistically significant differences were

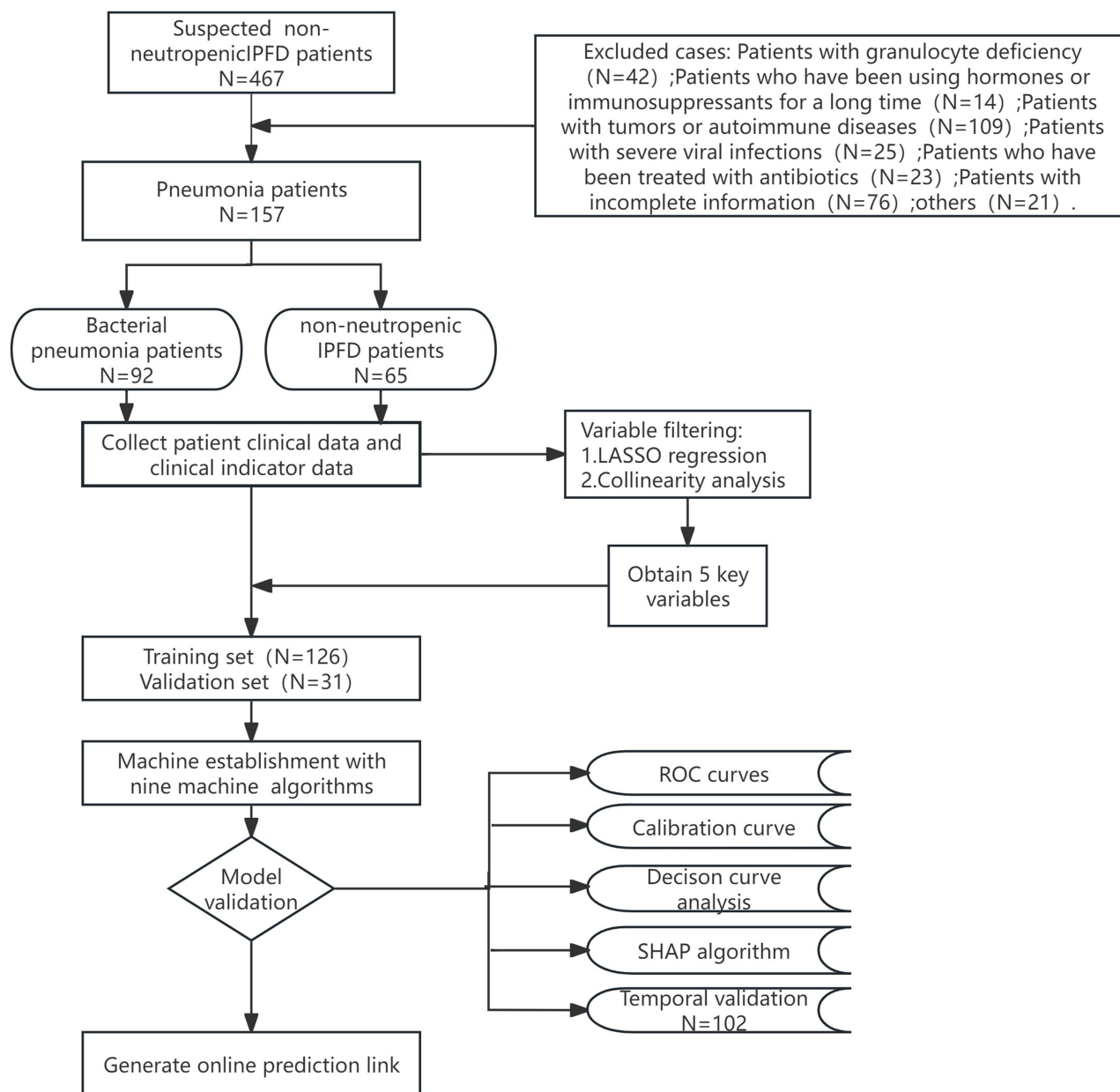


Figure 1 Flowchart of patient enrollment and study design. A total of 157 patients admitted between April 2018 and December 2022 were included in the primary cohort. An independent temporal validation cohort of 102 patients admitted between January 2023 and March 2025 was used to assess model generalizability. IPFD, invasive pulmonary fungal disease.

observed in demographic variables (gender, age) or other immunological markers, specifically CD3+ T cells, CD8+ T cells, NK cells, B cells, and mHLA-DR ($P > 0.05$). Detailed comparisons are summarized in [Table 1](#).

Dual Variable Selection in Fungal Diagnostic Modeling

To mitigate collinearity effects, variable selection was performed using LASSO regression with the “lambda.1se” regularization coefficient ($\lambda - 1$ standard error). The optimal λ value of 0.068 at 1 standard error from the minimum cross-validated error identified six initial predictors: BDG, GM, IL-6, mHLA-DR, monocyte count, and CD4+ T cells ([Figure 2](#)). Subsequent collinearity analysis using VIF was then applied to the six LASSO-selected predictors to ensure model stability. A VIF value quantifies how much the variance of a regression coefficient is inflated due to multicollinearity. A common rule of thumb is that a VIF > 10 indicates harmful multicollinearity that requires attention.¹³ This

Table 1 Baseline Demographic and Laboratory Characteristics of Patients with Bacterial Pneumonia and Non-Neutropenic Invasive Pulmonary Fungal Disease (IPFD)

Characteristic	Bacterial Group (n=92)	IPFD (n=65)	Statistic	P-Value
General information				
Male gender	60 (65.2)	39 (60.0)	$\chi^2 = 0.445$	0.505
Age (years)	68.000±15.820	66.769±14.704	$t = 0.745$	0.457
Laboratory tests				
BDG (pg/mL) ^a	60.183±52.197	106.483±118.717	$Z = -2.589$	0.007
GM (S/CO) ^a	0.240±0.502	0.724±1.068	$Z = -6.379$	<0.001
IL6 (pg/mL) ^a	156.391±290.846	429.449±822.942	$Z = -2.901$	0.004
CRP (mg/L) ^a	91.665±189.440	132.788±125.190	$Z = -3.852$	<0.001
CD3+T cells (%)	62.518±14.974	57.783±16.199	$t = 1.940$	0.053
CD4+T cells (%)	38.382±13.548	33.678±12.737	$t = 2.182$	0.031
CD8+T cells (%)	22.047±11.178	21.125±12.760	$t = 0.821$	0.412
NK cells (%)	20.041±12.874	22.695±14.783	$t = -1.025$	0.306
B cells (%)	17.586±12.948	18.118±13.659	$t = -0.184$	0.856
CD64 (%) ^a	2.034±3.974	6.220±22.188	$Z = -2.864$	0.004
mHLA-DR (%)	67.289±23.272	58.549±30.664	$t = 1.431$	0.153
Monocytes count (%)	3.670±3.066	2.600±1.942	$t = 2.639$	0.008

Notes: Data are presented as mean ± standard deviation or number (%). ^a Mann–Whitney *U*-test was used for analysis; all other continuous variables were analyzed using Student's *t*-test. Categorical data (Gender) were analyzed using the Chi-squared test.

Abbreviations: BDG, 1,3-β-D-glucan; GM, galactomannan; IL-6, interleukin-6; CRP, C-reactive protein; mHLA-DR, monocyte human leukocyte antigen-DR expression.

analysis excluded CD4+ T cells (VIF = 16.447), ultimately retaining five key predictors with acceptable multicollinearity levels (all VIF < 10): BDG, GM, IL-6, mHLA-DR, and monocyte count (Table 2).

Predictive Model Architecture Delineation

ROC curves were generated for each of the five selected predictors to evaluate their individual diagnostic performance. Among these variables, GM demonstrated the highest discriminative capacity with AUC of 0.730 (95% CI: 0.661–0.787), as illustrated in Figure 3A.

The dataset was partitioned into a training cohort (n=126) and validation cohort (n=31). These five optimized predictors were subsequently incorporated into nine machine learning algorithms for classification tasks. Through 5-fold cross-validation with AUC prioritization, the LightGBM model emerged as the optimal classifier in the validation set, achieving an AUC of 0.865 (95% CI: 0.728–0.999) and accuracy of 0.781 - surpassing individual biomarker performance. Model calibration analysis revealed strong agreement between predicted probabilities and observed outcomes (Figure 3B–D and Table 3). Decision curve analysis demonstrated clinical utility across the 0–1 probability threshold range, with sustained net benefit values (Figure 3E).

An independent test set (n=31) was randomly selected from the overall cohort, while the remaining 126 samples constituted the training set. The finalized LightGBM model maintained robust performance in the test set, yielding an AUC of 0.810 and accuracy of 0.750 (Figure 3F and G).

Finally, SHAP analysis was employed to interpret variable importance in the LightGBM model. The feature contribution plot demonstrated the following descending order of predictive impact: GM, mHLA-DR, monocyte count, IL-6, and BDG, as illustrated in Figure 3H.

Temporal Validation and Clinical Deployment of LightGBM Model

Temporal validation demonstrated robust diagnostic performance with an AUC of 0.821 (95% CI: 0.712–0.930), accuracy of 0.794, sensitivity of 0.667, specificity of 0.855, positive predictive value (PPV) of 0.688, negative predictive value (NPV) of 0.843, and F1-score of 0.677 (Figure 4). To enhance clinical applicability, we developed a user-friendly,

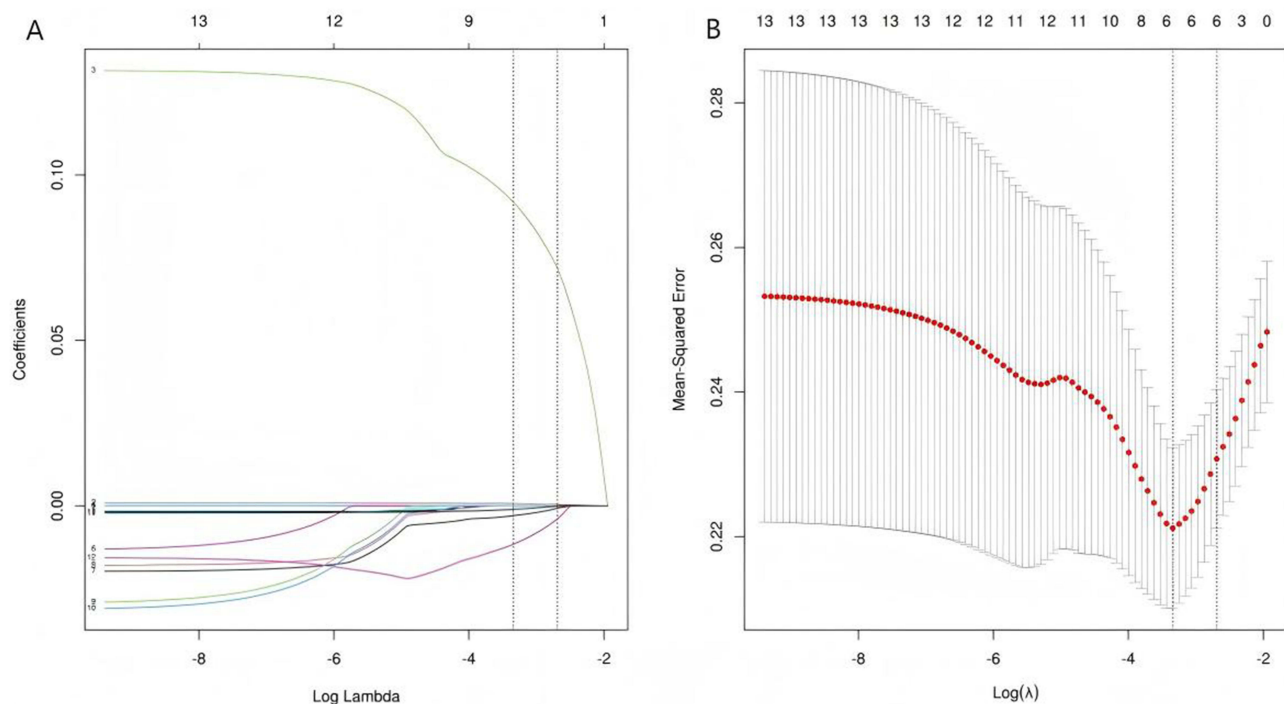


Figure 2 Feature selection using Least Absolute Shrinkage and Selection Operator (LASSO) regression. **(A)** LASSO coefficient profiles of the 12 candidate predictors. The vertical dashed line indicates the optimal value of the penalty parameter (λ) chosen by 10-fold cross-validation. **(B)** Plot of the cross-validation error (binomial deviance) versus $\log(\lambda)$. The left vertical dashed line indicates the λ value at which the model achieves minimum cross-validation error (λ_{\min}), and the right vertical dashed line indicates the largest λ value within one standard error of the minimum error (λ_{1se}), which was used for feature selection.

web-based prediction platform (<https://www.xsmartanalysis.com/model/list/predict/model/html?mid=27307andsymbol=117wD559BMzu531373vX>) that implements the optimized LightGBM algorithm. This tool allows clinicians to input routinely available biomarker values (GM, mHLA-DR, monocyte count, IL-6, BDG) and obtain real-time probability estimates for non-neutropenic IPFD, enabling rapid risk stratification and supporting clinical decision-making at the point of care. The use of commonly measured parameters ensures that the model can be integrated into existing clinical workflows without requiring additional specialized testing.

Discussion

The increasing incidence of non-neutropenic IPFD is closely linked to the widespread use of broad-spectrum antibiotics, corticosteroids, and immunosuppressants.¹⁸ Early diagnosis remains challenging due to nonspecific clinical manifestations and the invasiveness of histopathological confirmation, while conventional laboratory methods suffer from limited sensitivity and prolonged turnaround times.^{19,20} This underscores the clinical urgency to develop diagnostic models for timely intervention.

Several previous studies have attempted to develop diagnostic models for invasive fungal infections using various biomarkers and methodologies. For instance, Wang et al⁵ developed a deep learning model based on chest CT images for invasive pulmonary aspergillosis, while Cao et al⁷ created a risk prediction model using the MIMIC-IV database. Other studies have focused on single biomarkers or limited combinations, such as serum GM and BDG testing.^{21,22} However, few studies have specifically addressed the diagnostic challenges in non-neutropenic patients using a multimodal biomarker approach combined with machine learning algorithms.

Compared to these previous models, our LightGBM-based approach offers several distinct advantages: First, our model integrates both fungal antigens (GM, BDG) and host immune parameters (mHLA-DR, monocyte count, IL-6), providing a more comprehensive assessment of both pathogen presence and host response. Second, while many existing models require specialized imaging or complex laboratory tests, our model utilizes routinely available serological and immunological markers, making it more accessible for widespread clinical implementation. Third, our model

Table 2 Collinearity Analysis Using Variance Inflation Factor (VIF) for Variables Selected by LASSO Regression

Variable	VIF
CD3+T cells	52.720
NK cells	26.035
B cells	23.399
CD4 ⁺ T cells	16.447
CD8 ⁺ T cells	15.247
CD64	1.785
GM	1.721
Monocytes	1.318
IL-6	1.237
CRP	1.200
mHLADR	1.157
BDG	1.150
Age	1.095

Notes: VIF measures the severity of multicollinearity among the predictor variables. A VIF value > 10 is generally considered indicative of significant multicollinearity, warranting the exclusion of the variable from the model.

demonstrates superior performance (AUC: 0.865 in validation set) compared to single biomarkers or simpler models, with particularly improved sensitivity in non-neutropenic patients where traditional biomarkers often underperform.²¹ Finally, we have developed and deployed a user-friendly web-based prediction tool that facilitates real-time clinical decision support, enhancing practical utility beyond theoretical model development.

In this retrospective study involving 157 pneumonia patients (65 non-neutropenic IPFD cases vs 92 bacterial pneumonia controls), we identified five key predictors: GM, mHLA-DR, monocyte count, IL-6, and BDG, to construct nine machine learning models. Through 5-fold cross-validation, the LightGBM model demonstrated superior performance in the validation set (AUC: 0.865, 95% CI: 0.728–0.999; accuracy: 0.781) and test set (AUC: 0.810; accuracy: 0.750), outperforming individual biomarkers. Leveraging decision tree-based boosting, LightGBM efficiently integrates multimodal data while mitigating single-marker limitations.²³ Decision curve analysis revealed sustained net benefit across probability thresholds (0–1), with temporal validation further confirming generalizability (AUC: 0.821; accuracy: 0.794).

The selected biomarkers are clinically accessible: BDG and GM serve as cornerstone fungal diagnostics. However, serum GM exhibits variable sensitivity in non-neutropenic patients due to antigen clearance by intact immunity.²⁴ While Zhu et al²⁵ reported superior performance of metagenomic next-generation sequencing (mNGS) in bronchoalveolar lavage fluid compared to traditional GM testing, and Dai et al²¹ observed low GM/BDG sensitivity (22.2% and 9.4%, respectively) in non-neutropenic invasive pulmonary aspergillosis (IPA), our LightGBM model achieved markedly higher sensitivity (90.9% in training, 64.6% in validation) by synergizing fungal antigens with immune parameters. Notably, the model's specificity (90.8%) and positive predictive value (87.7%) surpassed those of serum GM alone (85.27% and 35.7%, respectively) in prior studies, underscoring the value of multimodal biomarker integration.²⁶

GM, released from fungal hyphae during invasion, aids early diagnosis but requires contextual interpretation.^{27,28} Host immunity critically shapes fungal infection dynamics,^{29–31} with monocyte-mediated fungal recognition via pattern recognition receptors (PRRs) driving antigen presentation and immune recruitment.^{32–34} Monocyte count and mHLA-DR levels thus complement fungal antigen detection by reflecting immune competence. IL-6 further enhances fungal clearance through immunocyte activation and antimicrobial peptide induction.^{35,36} BDG, a pan-fungal cell wall component, improves diagnostic sensitivity and enables therapeutic monitoring, particularly when serially assessed in ICU settings.^{22,37,38}

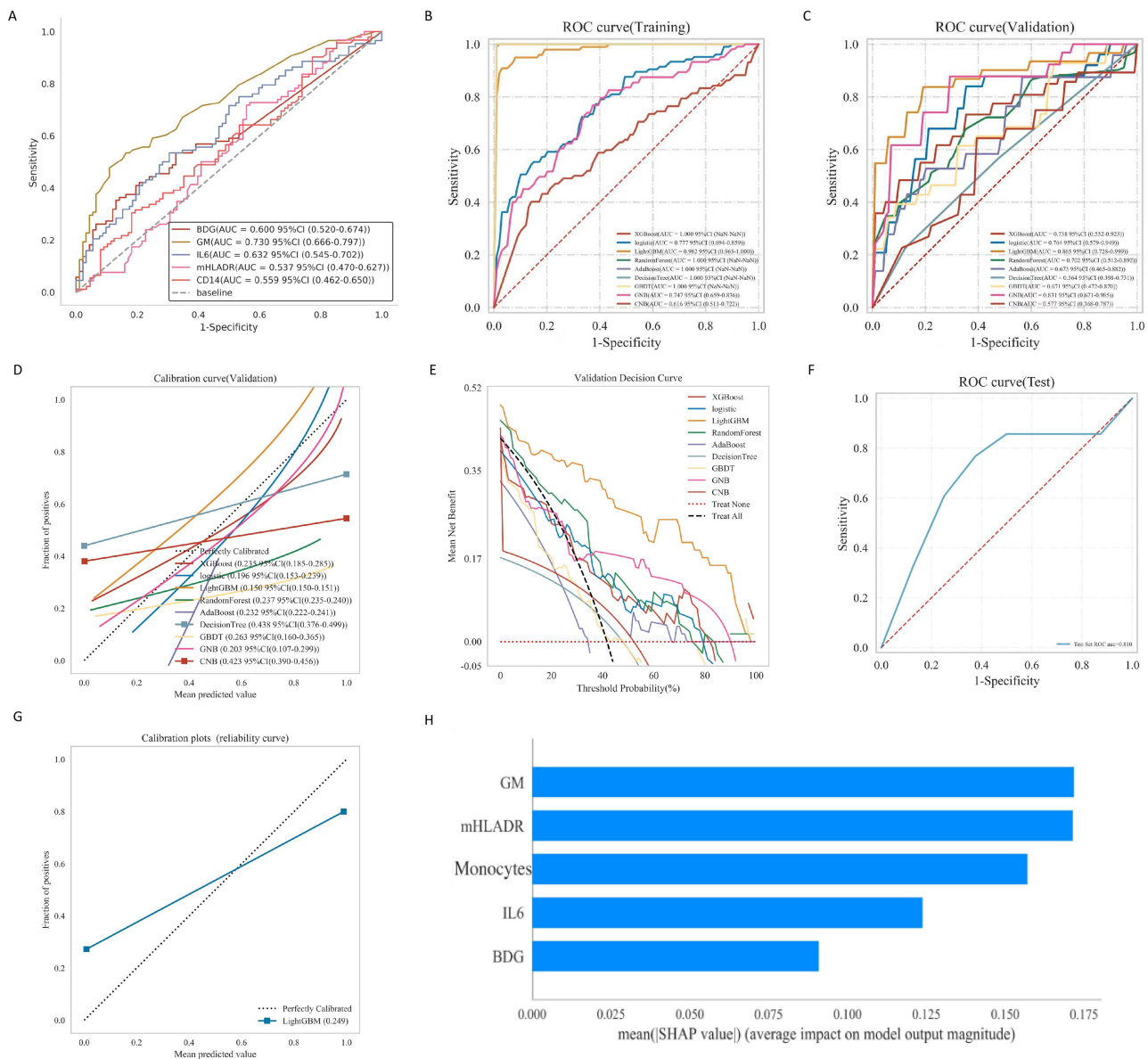


Figure 3 Performance evaluation of individual biomarkers and machine learning models. **(A)** Receiver operating characteristic (ROC) curves of the five selected biomarkers for discriminating non-neutropenic IPFD from bacterial pneumonia. **(B)** ROC curves of the nine machine learning models on the training set. **(C)** ROC curves of the models on the validation set. **(D)** Calibration curve for the LightGBM model on the validation set, showing the agreement between predicted probabilities and observed outcomes. The dashed line represents perfect calibration. **(E)** Decision curve analysis (DCA) for the LightGBM model in the validation cohort. The net benefit of using the model (solid black line) is compared against the strategies of treating all patients (solid grey line) and treating none patients (dashed black line). **(F)** ROC curve of the finalized LightGBM model on the independent test set. **(G)** Calibration curve of the LightGBM model on the test set. **(H)** Summary plot of Shapley Additive exPlanations (SHAP) values for the LightGBM model, illustrating the mean impact of each feature on the model output. The features are ordered by their mean absolute SHAP value, representing overall importance.

Regarding the practical applicability of the model, all five biomarkers (GM, BDG, mHLA-DR, monocyte count, and IL-6) are routinely measured in most clinical laboratories using standardized assays (eg, ELISA for GM and BDG, flow cytometry for mHLA-DR and monocyte count, and immunoassays for IL-6). The implementation cost is relatively low, as these tests are already widely available and reimbursed in many healthcare systems. The integration of these parameters into a single predictive model does not require additional specialized equipment or training, making it feasible for adoption in diverse clinical settings, including resource-limited hospitals. The web-based prediction tool further reduces operational barriers by providing an intuitive, freely accessible platform for real-time risk calculation.

Table 3 Comparative Performance Metrics of Nine Machine Learning Algorithms for Discriminating Non-Neutropenic IPFD From Bacterial Pneumonia in the Training and Validation Sets

Models	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score
Training set							
XGBoost	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Logistic	0.777	0.708	0.742	0.682	0.656	0.790	0.684
LightGBM	0.982	0.952	0.909	0.980	0.968	0.943	0.937
RandomForest	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AdaBoost	1.000	0.996	1.000	0.993	0.991	1.000	0.996
DecisionTree	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GBDT	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GNB	0.747	0.700	0.796	0.632	0.603	0.815	0.686
CNB	0.616	0.680	0.431	0.851	0.669	0.685	0.524
Validation set							
XGBoost	0.738	0.688	0.350	0.950	0.875	0.668	0.467
Logistic	0.764	0.688	0.647	0.721	0.611	0.776	0.606
LightGBM	0.865	0.781	0.646	0.908	0.871	0.731	0.742
RandomForest	0.702	0.625	0.583	0.693	0.625	0.650	0.586
AdaBoost	0.673	0.672	0.528	0.748	0.508	0.764	0.511
DecisionTree	0.564	0.562	0.429	0.699	0.545	0.624	0.440
GBDT	0.671	0.672	0.429	0.821	0.590	0.700	0.495
GNB	0.831	0.688	0.847	0.597	0.616	0.833	0.697
CNB	0.577	0.547	0.393	0.667	0.477	0.586	0.431

Moreover, the model's ability to provide rapid, non-invasive diagnostic support may help reduce the need for costly and invasive procedures (eg, bronchoscopy or lung biopsy), shorten time to appropriate antifungal therapy, and potentially improve patient outcomes. Future cost-effectiveness analyses are warranted to further validate the economic benefits of model implementation.

Our web-based prediction tool (<http://www.xsmartanalysis.com/>) leverages these five routine biomarkers to enable real-time, cost-effective risk stratification for non-neutropenic IPFD, reducing diagnostic barriers in primary care.

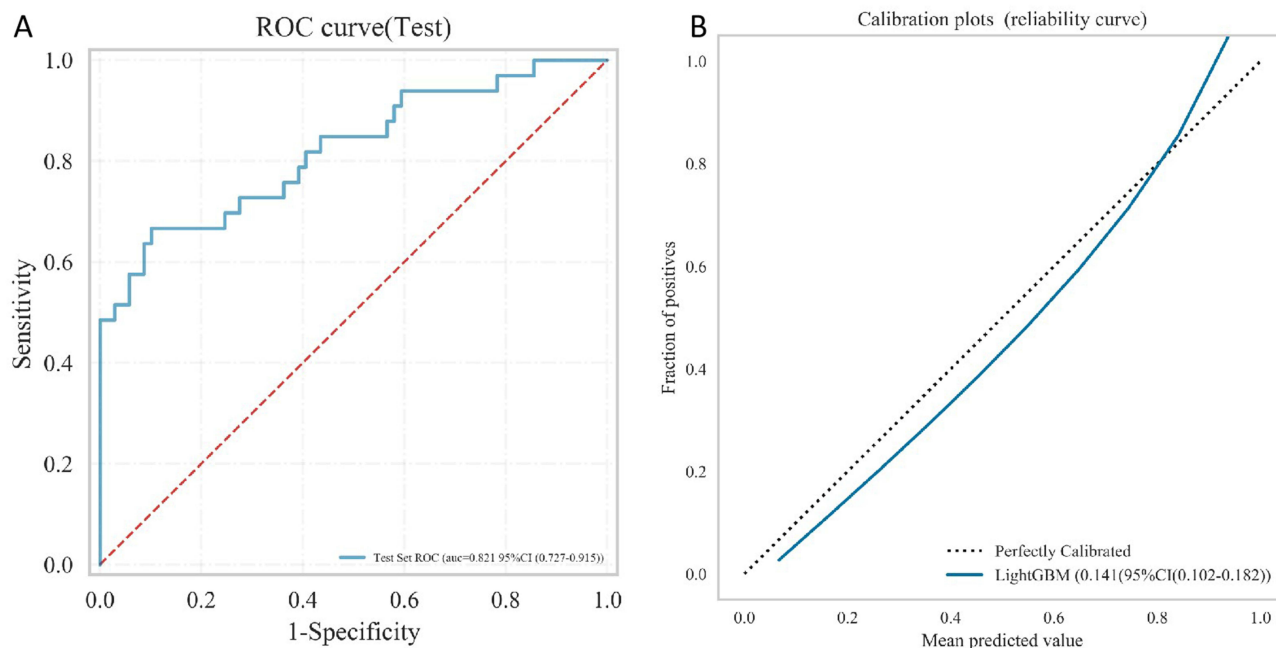


Figure 4 Temporal validation Performance of the LightGBM Diagnostic Model. **(A)** Receiver operating characteristic (ROC) curve of the model applied to the independent temporal validation cohort (January 2023 - March 2025). **(B)** Calibration curve showing the relationship between predicted probability of non-neutropenic IPFD and the actual observed frequency in the temporal validation set.

Limitations include the single-center retrospective design and modest sample size, which may affect the generalizability of our findings. Furthermore, the intentional exclusion of patients with certain comorbidities (recent steroid use, malignancies, and severe viral infections) was necessary to establish a foundational model in a less confounded population but limits immediate applicability to these important patient subgroups where non-neutropenic IPFD frequently occurs. Additionally, while the selected biomarkers are clinically available, the requirement for specialized immunologic testing (particularly mHLA-DR quantification) may present implementation challenges in resource-limited settings. Multicenter validation with expanded cohorts that include patients with these comorbidities is warranted to refine model robustness and enhance its generalizability.

Conclusions

We developed a LightGBM-based diagnostic model integrating GM, BDG, mHLA-DR, monocyte count, and IL-6, achieving AUC of 0.865 (validation) and 0.821 (external testing). The accompanying web calculator facilitates rapid differentiation of non-neutropenic IPFD from bacterial pneumonia, offering a practical solution for early diagnosis. Future multicenter studies will further validate its clinical utility.

Abbreviations

AUC, Area Under the Receiver Operating Characteristic Curve; BDG, 1,3- β -D-glucan; BMI, Body Mass Index; CRP, C-reactive Protein; DCA, Decision Curve Analysis; ELISA, Enzyme-Linked Immunosorbent Assay; ECLIA Electrochemiluminescence immunoassay; GM, Galactomannan; H&E, Hematoxylin and Eosin; HLA-DR, Human Leukocyte Antigen-DR; IL-6, Interleukin-6; IPFD, Invasive Pulmonary Fungal Disease; LASSO, Least Absolute Shrinkage and Selection Operator; mHLA-DR, Monocyte HLA-DR Expression; MIMIC-IV, Medical Information Mart for Intensive Care IV; NPV, Negative Predictive Value; PPV, Positive Predictive Value; ROC, Receiver Operating Characteristic Curve; SHAP, Shapley Additive exPlanations.

Data Sharing Statement

The data that support the findings of this study are available from the corresponding authors (email: juyua1989@163.com) upon reasonable request.

Ethics Approval and Consent to Participate

This study received ethical approval from the Ethics Committee of the First Affiliated Hospital of Guangzhou University of Chinese Medicine (NO. K-2024-177), with procedural adherence to the ethical guidelines established by the World Medical Association's Declaration of Helsinki. The requirement for written informed consent was deemed unnecessary by the Ethics Committee owing to the retrospective nature of the study, which utilized anonymized patient records exclusively.

Acknowledgments

This work is supported by the the Beckman Coulter DxAI 800 analytical system.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors declare no competing interests in this work.

References

- Denning DW. Global incidence and mortality of severe fungal disease. *Lancet Infect Dis.* 2024;24(7):e428–e438. doi:10.1016/S1473-3099(23)00692-8
- Bassetti M, Giacobbe DR, Agvald-Ohman C, et al. Invasive Fungal Diseases in Adult Patients in Intensive Care Unit (FUNDICU): 2024 consensus definitions from ESGCIP, EFISG, ESICM, ECOMM, MSGERC, ISAC, and ISHAM. *Intensive Care Med.* 2024;50(4):502–515. doi:10.1007/s00134-024-07341-7
- Bongomin F, Gago S, Oladele RO, Denning DW. Global and Multi-National Prevalence of Fungal Diseases-Estimate Precision. *J Fungi.* 2017;3(4):1.
- Azim A, Ahmed A. Diagnosis and management of invasive fungal diseases in non-neutropenic ICU patients, with focus on candidiasis and aspergillosis: a comprehensive review. *Front Cell Infect Microbiol.* 2024;14:1256158. doi:10.3389/fcimb.2024.1256158
- Wang W, Li M, Fan P, et al. Prototype early diagnostic model for invasive pulmonary aspergillosis based on deep learning and big data training. *Mycoses.* 2023;66(2):118–127. doi:10.1111/myc.13540
- Zhang K, Zhao G, Liu Y, et al. Clinic, CT radiomics, and deep learning combined model for the prediction of invasive pulmonary aspergillosis. *BMC Med Imaging.* 2024;24(1):264. doi:10.1186/s12880-024-01442-x
- Cao Y, Li Y, Wang M, et al. Interpretable Machine Learning For Predicting Risk Of Invasive Fungal Infection In Critically Ill Patients In The Intensive Care Unit: a Retrospective Cohort Study Based On MIMIC-IV Database. *Shock.* 2024;61(6):817–827. doi:10.1097/SHK.0000000000002312
- Ye F, Zeng P, Li Z, et al. Detection of Aspergillus DNA in BALF by Real-time PCR and Galactomannan Antigen for the Early Diagnosis of Chronic Pulmonary Aspergillosis. *Ann Clin Lab Sci.* 2021;51(5):698–704.
- Lu Y, Liu L, Li H, et al. The clinical value of Aspergillus-specific IgG antibody test in the diagnosis of nonneutropenic invasive pulmonary aspergillosis. *Clin Microbiol Infect.* 2023;29(6):797.e791–797.e797. doi:10.1016/j.cmi.2023.02.002
- Wang H, Chen X, You H, et al. Performance of multiplex PCR-based targeted next-generation sequencing in bronchoalveolar lavage fluid for the diagnosis of invasive pulmonary aspergillosis in non-neutropenic patients. *J Infect Dis.* 2025;231(6):1609–1618. doi:10.1093/infdis/jiaf044
- Donnelly JP, Chen SC, Kauffman CA, et al. Revision and Update of the Consensus Definitions of Invasive Fungal Disease From the European Organization for Research and Treatment of Cancer and the Mycoses Study Group Education and Research Consortium. *Clin Infect Dis.* 2020;71(6):1367–1376. doi:10.1093/cid/ciz1008
- Tibshirani R, Tibshirani R. *Regression Shrinkage via the Lasso.* 1996.
- James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: An Introduction to Statistical Learning.* 2013.
- Vickers AJ, Elkin EB. Decision Curve Analysis: a Novel Method for Evaluating Prediction Models. *Med Decis Mak.* 2006;26(6):565–574. doi:10.1177/0272989X06295361
- Vickers AJ, Calster BV, Steyerberg EW, Krone R, Brown DL. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic Prognostic Res.* 2019;3:3. doi:10.1186/s41512-019-0048-7
- Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. 2017.
- Steyerberg EW. Clinical Prediction Models: a Practical Approach to Development, Validation, and Updating by Ewout W. Steyerberg. *J R Stat Soc.* 2010;66(2):661–662.
- Sun C, Cai X, Zhong H, et al. Pentraxin-3 as a novel prognostic biomarker in non-neutropenic invasive pulmonary aspergillosis patients. *Microbiol Spectr.* 2025;13(3):e0294524. doi:10.1128/spectrum.02945-24
- Wang H, Yu D, Chen X, et al. Performance of rapid on-site evaluation of touch imprints of bronchoscopic biopsies or lung tissue biopsies for the diagnosis of invasive pulmonary filamentous fungi infections in non-neutropenic patients. *J Clin Microbiol.* 2024;62(7):e0047924. doi:10.1128/jcm.00479-24
- Barros N, Rosenblatt RE, Phipps MM, Fomin V, Mansour MK. Invasive fungal infections in liver diseases. *Hepatol Commun.* 2023;7(9). doi:10.1097/HCP.0000000000000216
- Dai Z, Cai M, Yao Y, et al. Comparing the diagnostic value of bronchoalveolar lavage fluid galactomannan, serum galactomannan, and serum 1,3-β-d-glucan in non-neutropenic respiratory disease patients with invasive pulmonary aspergillosis. *Medicine.* 2021;100(14):e25233. doi:10.1097/MD.00000000000025233
- Wu Z, Wang L, Tan L, Wu J, Chen Z, Hu M. Diagnostic value of galactomannan in serum and bronchoalveolar lavage fluid for invasive pulmonary aspergillosis in non-neutropenic patients. *Diagn Microbiol Infect Dis.* 2021;99(4):115274. doi:10.1016/j.diagmicrobio.2020.115274
- Sun B, Lei M, Wang L, et al. Prediction of sepsis among patients with major trauma using artificial intelligence: a multicenter validated cohort study. *Int J Surg.* 2025;111(1):467–480. doi:10.1097/JS9.0000000000001866
- Lim SY, Lee YW, Jung J, et al. Diagnostic yield of a bronchoalveolar lavage fluid galactomannan assay in patients with negative serum galactomannan results suspected to have invasive pulmonary aspergillosis. *Mycoses.* 2021;64(9):1124–1131. doi:10.1111/myc.13269
- Zhu N, Zhou D, Xiong W, Zhang X, Li S. Performance of mNGS in bronchoalveolar lavage fluid for the diagnosis of invasive pulmonary aspergillosis in non-neutropenic patients. *Front Cell Infect Microbiol.* 2023;13:1271853. doi:10.3389/fcimb.2023.1271853
- Chen F, Chen Y, Chi Y, Gao T, Zhao Y, Shao H. Diagnosis of invasive pulmonary fungal infections by a real-time panfungal PCR assay in non-neutropenic patients. *Medicine.* 2023;102(51):e36385. doi:10.1097/MD.00000000000036385
- Nuh A, Ramadan N, Shah A, Armstrong-James D. Sputum Galactomannan Has Utility in the Diagnosis of Chronic Pulmonary Aspergillosis. *J Fungi.* 2022;8(2):1.
- Schub T, Klugherz I, Wagener J, et al. Serum antigen tests for the diagnosis of invasive aspergillosis: a retrospective comparison of five Aspergillus antigen assays and one beta-D-glucan assay. *J Clin Microbiol.* 2024;62(12):e0095024. doi:10.1128/jcm.00950-24
- He Q, Cao J, Zhang M, Feng C. IL-17 in plasma and bronchoalveolar lavage fluid in non-neutropenic patients with invasive pulmonary aspergillosis. *Front Cell Infect Microbiol.* 2024;14:1402888. doi:10.3389/fcimb.2024.1402888

30. Heung LJ, Wiesner DL, Wang K, Rivera A, Hohl TM. Immunity to fungi in the lung. *Semin Immunol.* 2023;66:101728. doi:10.1016/j.smim.2023.101728
31. Jaggi TK, Agarwal R, Tiew PY, et al. Fungal lung disease. *Eur Respir J.* 2024;64(5):2400803. doi:10.1183/13993003.00803-2024
32. Jannuzzi GP, de Almeida JRF, Paulo LNM, de Almeida SR, Ferreira KS. Intracellular PRRs Activation in Targeting the Immune Response Against Fungal Infections. *Front Cell Infect Microbiol.* 2020;10:591970. doi:10.3389/fcimb.2020.591970
33. Weerasinghe H, Stölting H, Rose AJ, Traven A. Metabolic homeostasis in fungal infections from the perspective of pathogens, immune cells, and whole-body systems. *Microbiol Mol Biol Rev.* 2024;88(3):e0017122. doi:10.1128/mmbr.00171-22
34. Loh JT, Lam KP. Fungal infections: immune defense, immunotherapies and vaccines. *Adv Drug Deliv Rev.* 2023;196:114775. doi:10.1016/j.addr.2023.114775
35. Liu F, Zhang X, Du W, et al. Diagnosis values of IL-6 and IL-8 levels in serum and bronchoalveolar lavage fluid for invasive pulmonary aspergillosis in chronic obstructive pulmonary disease. *J Investig Med.* 2021;69(7):1344–1349. doi:10.1136/jim-2021-001857
36. Shankar J, Thakur R, Clemons KV, Stevens DA. Interplay of Cytokines and Chemokines in Aspergillosis. *J Fungi.* 2024;10(4). doi:10.3390/jof10040251
37. Friedman DZP, Schwartz IS. Emerging Diagnostics and Therapeutics for Invasive Fungal Infections. *Infect Dis Clin North Am.* 2023;37(3):593–616. doi:10.1016/j.idc.2023.05.001
38. Scharmann U, Verhasselt HL, Kirchhoff L, Furnica DT, Steinmann J, Rath PM. Microbiological Non-Culture-Based Methods for Diagnosing Invasive Pulmonary Aspergillosis in ICU Patients. *Diagnostics.* 2023;13(16). doi:10.3390/diagnostics13162718

Infection and Drug Resistance

Publish your work in this journal

Infection and Drug Resistance is an international, peer-reviewed open-access journal that focuses on the optimal treatment of infection (bacterial, fungal and viral) and the development and institution of preventive strategies to minimize the development and spread of resistance. The journal is specifically concerned with the epidemiology of antibiotic resistance and the mechanisms of resistance development and diffusion in both hospitals and the community. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/infection-and-drug-resistance-journal>

Dovepress
Taylor & Francis Group