

**Characteristics of Gut Microbiota in Patients with Severe Pneumonia
and its Potential Clinical Relevance**

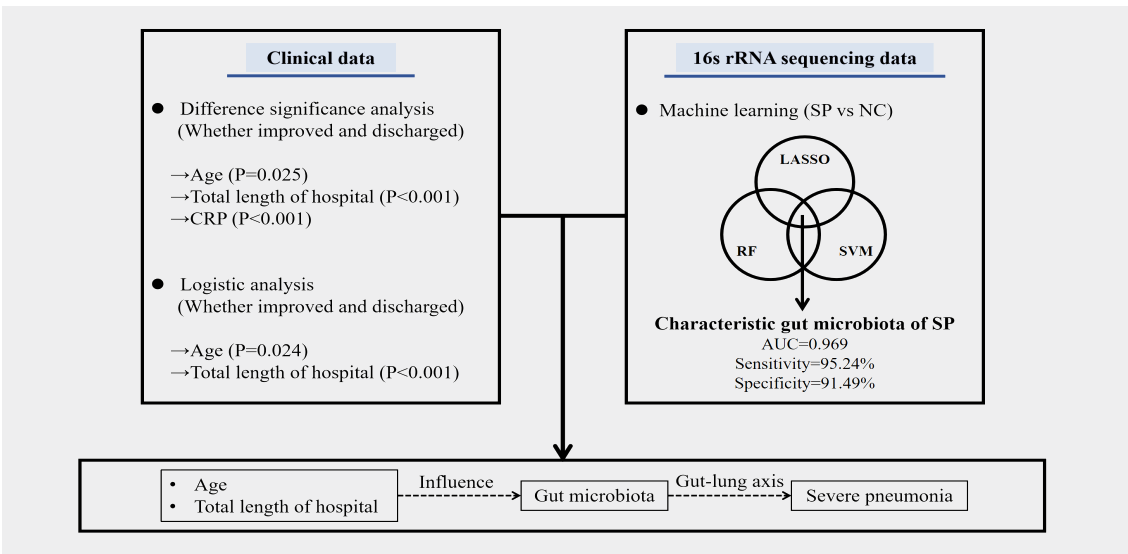
FANGCHAO ZHONG¹, MAOSEN HUANG¹, XIAOXIA WEI¹,
LIHUA FU¹ and LINHAI YAN^{1*}

¹Department of Gastrointestinal Surgery, Guangxi Medical University Cancer Hospital,
Nanning, Guangxi Zhuang Autonomous Region, China

Submitted 5 June 2025, accepted 5 September 2025, published online 31 July 2026

Abstract

Gut microbiota is associated with a variety of diseases, but its relationship with severe pneumonia remains to be explored. This study primarily analyzed the intestinal microbiota of patients with severe pneumonia and examined its association with clinical data. We collected clinical data from 96 patients with severe pneumonia for differential analysis and identified prognostic factors using logistic regression. Fecal samples from patients with severe pneumonia and healthy controls were collected and analyzed using 16S rRNA sequencing. We applied three machine learning algorithms (LASSO, Random Forest, and SVM) to identify microbial markers associated with severe pneumonia. The patients were grouped by “discharge status”. Significant differences were observed in age ($p = 0.025$), total length of hospital stay ($p < 0.001$), and C-reactive protein (CRP) ($p < 0.001$). Logistic regression analysis identified age ($p = 0.024$) and total hospital stay ($p < 0.001$) as factors influencing the likelihood of improvement and discharge. Diversity analysis of collected stool samples revealed differences between the two groups. LDA Effect Size (LefSe) analysis highlighted significant microbial differences at various taxonomic levels between the two populations. Three machine learning algorithms identified 9 microbial markers for severe pneumonia. A diagnostic prediction model was constructed, with an area under the Receiver Operating Characteristic (ROC) curve of 0.969 (95% CI: 0.946–0.992). Patients with severe pneumonia exhibit unique intestinal microbiota characteristics, which may be regulated by age and total length of hospital stay, thereby influencing the disease’s prognosis.



Key words: severe pneumonia; gut microbiota; 16sRNA sequencing; machine learning

* Corresponding author: L. Yan, Department of Gastrointestinal Surgery, Guangxi Medical University Cancer Hospital, Nanning, Guangxi Zhuang Autonomous Region, China; [email: yanlinhai000@163.com](mailto:yanlinhai000@163.com)

© 2026 Fangchao Zhong et al.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Severe pneumonia (SP) is a critical pulmonary infection caused by various pathogens, often accompanied by complications such as acute respiratory failure and multiple organ dysfunction, which consequently lead to high mortality and disability risks (Torres et al. 2021). In recent years, researchers have been consistently investigating the diverse factors influencing lung infections in clinical practice, with particular focus on risk factors associated with malnutrition in pneumonia patients. Findings have indicated that factors such as obesity and advanced age exert a significant influence on the susceptibility to severe pneumonia (Silverio et al. 2021). Elucidating the specific pathways through which these factors influence the pathogenesis and progression of pneumonia remains a question worthy of in-depth exploration. Previous investigations have demonstrated that host-related factors, such as obesity and age, can modulate the human microbiota, disrupt host immune homeostasis, and thereby contribute to disease development (Honarpisheh et al. 2022). The human body harbors a microecosystem characterized by a diverse composition, in which the microbiota, comprising bacteria, viruses, and fungi, inhabits various anatomic sites throughout the body, notably the lungs and gut (Hashimoto 2023).

The lung microbiota has been found to be associated with a substantial number of pulmonary diseases. Tian et al. (2024) recognized that respiratory tract microbiota-based interventions hold promise for revolutionizing asthma management. Budden and Romani (2025) through an analysis of the regulatory roles of respiratory tract microbiota in pneumonia, lung cancer, and chronic obstructive pulmonary disease (COPD), further concluded that microbiota-targeted therapeutic approaches are feasible. These studies suggest that the lung microbiome may modulate pulmonary inflammation; however, the specific mechanisms underlying this remain to be elucidated. Additionally, whether this microbiota is associated with the gut microbiota remains inconclusive.

Compared to the lung microbiome, the gut microbiota has long been the focus of extensive research due to its high relative abundance and species diversity, and has been more thoroughly investigated to elucidate its mechanistic roles. Accumulating evidence demonstrates that gut microbiota dysbiosis is closely associated with the onset and progression of diverse patho-

logical conditions, including common gastrointestinal disorders such as inflammatory bowel disease (Qiu et al. 2022) and colorectal cancer (Wong and Yu 2023), as well as systemic diseases such as obesity (Gomes et al. 2018), diabetes (Iatcu et al. 2021), and rheumatoid arthritis (Zhao et al. 2022).

Studies have demonstrated that the use of composite probiotic formulations in the treatment of pneumonia patients can effectively improve prognosis and accelerate recovery (Gutiérrez-Castrellón et al. 2022). With the identification of the impact of gut health on pulmonary immunity, the concept of the “gut-lung axis” has garnered significant attention. Intestinal microbiota exert both local effects within the gastrointestinal tract and distal influences on distant organs, such as the lungs. This bidirectional interaction, termed the “gut-lung axis”, serves as the foundational mechanism underlying the use of the microbiome in the prevention or treatment of respiratory diseases (Bulanda and Wypych 2022). In recent years, dysbiosis of the gut microbiota has been considered to play a key role in the occurrence, progression, and prognosis of pneumonia (Nagata et al. 2023). These findings provide important insights for further exploration of the mechanisms of the gut-lung axis. Patients with pulmonary diseases are frequently associated with dysbiosis, which is characterized by the over proliferation of pathogenic bacteria and a reduction in beneficial bacteria. This dysbiotic state may exacerbate pulmonary inflammatory responses and potentially influence clinical outcomes (Li et al. 2024). In severe pulmonary infections, alterations in the gut microbiota are recognized as a critical factor contributing to immune dysfunction, excessive inflammatory activation, and a poor prognosis (Yang and Cong 2021). The gut and lung engage in bidirectional interactions via multiple mechanisms, including circulating inflammatory cells and mediators, the intestinal barrier, oxidative stress, immune responses, and other related processes (Wang et al. 2023).

The current research is increasingly focusing on exploring the relationship between the gut microbiota and pneumonia. However, studies on the gut microbiota in patients with severe pneumonia remain relatively limited. This study aims to analyze gut microbiota characteristics in patients with severe pneumonia using a machine learning algorithm and to discuss potential correlations between gut microbiota and clinical prognosis in this population.

Experimental

Materials and Methods

Study population. Severe pneumonia patients diagnosed at Rongshui County People's Hospital between September 2022 and May 2023 were selected, and 105 volunteers were included as healthy controls. Fecal samples and clinical data from the study population were then collected. Inclusion criteria: i) The disease group should meet the diagnostic criteria for severe pneumonia, as detailed in Table S1; ii) The age of the host in each group is over 18 years old; iii) The clinical data of the disease group is complete, and all groups have signed the consent form. Exclusion criteria: i) Malignant tumors; ii) Pregnant or lactating women; iii) History of immunodeficiency; iv) History of significant disease. After screening, 96 patients with severe pneumonia and 105 volunteers were included for fecal intestinal microbiota analysis. This study has obtained ethical approval from the Medical Ethics Committee of Rongshui Miao Autonomous County People's Hospital (approval number: rsry-II-2025-03) and has been strictly conducted in accordance with the principles outlined in the Declaration of Helsinki.

Intestinal microbiota sequencing. Morning fecal samples (10 g) were collected from patients with severe pneumonia and healthy controls enrolled in the study. Total bacterial genomic DNA samples were extracted, and the quantity and quality of the extracted DNA were measured, respectively.

The V4 region of the bacterial 16S rRNA gene was amplified via PCR. Library construction was performed using the TruSeq[®] DNA PCR-Free library preparation kit (Illumina, Inc., USA). Then, Qubit quantification was performed to assess library quality, and only libraries that passed quality control were sequenced on a NovaSeq[™] 6000 platform (Illumina, Inc., USA).

Raw sequencing reads with exact matches to the barcodes were assigned to respective samples and identified as valid sequences. The low-quality sequences were filtered (Gill et al. 2006; Chen and Jiang 2014). Paired-end reads were assembled, then Operational Taxonomic Unit (OTU) picking using Vsearch v2.15.0 (Rognes et al. 2016). The Quantitative Insights Into Microbial Ecology QIIME2 v2020.6 (Bolyen et al. 2019) pipeline was employed to process the OTUs and sequencing data. OTU taxonomic classification was conducted by searching the representative sequences set against the SILVA138 database (Quast et al. 2013).

An OTU table was generated to record the abundance of each OTU in each sample and its taxonomy.

Subsequently, data clustering was performed at the OTU level. Taxonomic annotation was conducted for representative sequences of each OTU, followed by evaluation of species richness and diversity within the samples. Alpha and beta diversity analyses were further carried out for the gut microbiota to assess differences in microbial community composition and structure across samples (The detailed version of the intestinal microbiota sequencing can be found in Supplementary Information S1).

Machine learning method. In this study, we utilized R (v4.4.1; R Core Team 2024) for LASSO machine learning analysis, random forest (RF) analysis, and support vector machine (SVM) analysis to identify the core intestinal microbiota associated with severe pneumonia patients. First, we used the glmnet package for LASSO regression, setting the random seed to "123" and employing mean squared error as the model evaluation metric. The optimal λ value (the minimum λ value) was determined using 5-fold cross-validation, yielding 0.073, corresponding to the best model. Next, the LASSO model was refitted, and the coefficients corresponding to the minimum λ value were identified by observing changes in the coefficients across different λ values. The indices of the non-zero coefficients were identified, and the functional bacterial genera were screened. Finally, the names of the bacterial genera with non-zero coefficients and their corresponding coefficient values were extracted. Subsequently, the randomForest package was used for RF analysis to assess the importance of bacterial genera and identify the bacteria with the most tremendous impact. Next, SVM analysis was conducted using the e1071 package to identify the core bacterial genera based on this analytical approach. Finally, by intersecting the results from LASSO, RF, and SVM analyses, the core bacterial genera most strongly associated with SP were identified, providing a foundation for subsequent biomarker studies.

Statistical methods. All data were analyzed by IBM SPSS Statistics for Windows v27.0 (IBM Corp., USA), R language v4.4.1, GraphPad Prism 9 software (www.graphpad.com). The mean value $\pm$ standard deviation (Mean $\pm$ SD) was used to express the data, and the *t*-test or Mann-Whitney *U* test was used for inter-group comparison; the frequency and percentage were used to express the data, and the chi-square test was used for inter-group comparison. A *p*-value < 0.05 was considered statistically significant.

Results

Clinical characteristics of the severe pneumonia population. In this study, 96 patients with SP were included for clinical data analysis. Among them, 45 patients were discharged after improvement, and 51 patients were discharged without improvement. According to the statistical analysis of clinical data differences, it can be seen from Table I that the age of patients with SP in the improved discharge group was

significantly lower than that in the non-improved discharge group ($p = 0.025$). The total length of stay of patients in the improved discharge group was significantly longer than that of patients in the non-improved discharge group ($p < 0.001$). In addition, the C-reactive protein (CRP) levels of patients in the non-improved discharge group were significantly elevated ($p < 0.001$). When the body is stimulated by infections, inflammation, or tissue damage, CRP in the blood will rise rapidly. This accurately reflects the poor discharge status

Table I
Comparison of clinical data between the improved discharge group and the non-improved discharge group.

Item	Improved discharge group (n = 45)	Non-improved discharge group (n = 51)	t/ χ^2	p
Gender			0.435	0.510
Male	28	35		
Female	17	16		
Age	64.577 ± 15.878	71.804 ± 15.277	−2.270	0.025
Total length of hospital stay	14.533 ± 4.283	7.686 ± 6.351	6.108	< 0.001
APACHE II score	24.133 ± 5.880	26.520 ± 7.393	−1.728	0.087
Past history				
Hypertension			0.996	0.318
Yes	21	29		
No	24	22		
Diabetes			2.438	0.118
Yes	13	8		
No	32	43		
Coronary disease			0.269	0.604
Yes	7	10		
No	38	41		
Personal history				
Smoke			0.041	0.840
Yes	15	18		
No	30	33		
Excessive drinking			1.226	0.268
Yes	12	19		
No	33	32		
Test results				
WBC (×10 ⁹ /l)	14.706 ± 6.293	14.020 ± 9.545	0.402	0.689
N (%)	85.491 ± 9.472	82.269 ± 17.070	1.100	0.274
HGB (g/l)	98.591 ± 40.885	99.522 ± 35.780	−0.117	0.907
PLT (×10 ⁹ /l)	224.767 ± 116.618	201.880 ± 133.020	0.875	0.384
CRP (mg/l)	61.623 ± 55.878	128.030 ± 75.944	−4.293	< 0.001
INR	1.196 ± 0.436	1.241 ± 0.442	−0.493	0.623
Alveolar lavage fluid				
Gram-Negative bacillus			0.178	0.673
(+)	24	25		
(−)	21	26		

of patients who did not improve. We then proceeded to use “whether the patient was discharged after improvement” as the dependent variable, with clinical characteristic indicators related to severe pneumonia patients, such as gender, age, total length of stay, APACHE II score, past history (hypertension, diabetes, coronary disease), personal history (smoke, excessive drinking), and relevant test results, as independent variables. Logistic regression analysis was performed, and the assignment details are shown in Table II.

Through univariate and multivariate logistic regression analyses, we identified age ($p = 0.031$) and total hospital stay ($p < 0.0001$) as significant factors influencing the outcome in patients with severe pneumonia. As shown in the results of Table III, increasing age (OR = 0.970, 95% CI: 0.942–0.996) was found to be a risk factor for discharge after improvement in severe pneumonia patients, while a more extended total hospital stay (OR = 1.255, 95% CI: 1.147–1.398) was identified as a protective factor for discharge after improvement.

Table II
Assignment method.

Variable	Assignment
Dependent variable	
Whether the patient was discharged after improvement	No = 0, Yes = 1
Independent variable	
Gender	Female = 0, Male = 1
Age	Substitute the original value
Total length of stay	Substitute the original value
APACHE II score	Substitute the original value
Hypertension	No = 0, Yes = 1
Diabetes	No = 0, Yes = 1
Coronary disease	No = 0, Yes = 1
Smoke	No = 0, Yes = 1
Excessive drinking	No = 0, Yes = 1
WBC ($\times 10^9/l$)	Substitute the original value
HGB (g/l)	Substitute the original value
PLT ($\times 10^9/l$)	Substitute the original value
CRP (mg/l)	Substitute the original value
Gram-Negative bacillus	“−” = 0, “+” = 1

Analysis of bacterial characteristics in severe pneumonia patients. To compare gut microbiota between patients with severe pneumonia and healthy controls, this study collected fecal samples from both groups and analyzed microbial community diversity using 16S rRNA sequencing. Statistical analysis showed no significant differences in age or gender between the two groups (Table SII). Fig. 1A–1F presents the α -diversity analysis of patients with severe pneumonia, where the Richness, Chao 1, and Observed Species indices were significantly higher than those in the healthy control group ($p = 0.017$). These indices suggest potential alterations in the gut microbiota of patients with severe pneumonia. Subsequent principal component analysis (PCA) and principal coordinate analysis (PCoA) revealed partial differences in the clustered distribution patterns of microbial

communities between the disease and control groups, further indicating potential discrepancies in microbial community distribution between the two groups (Fig. 1G–1H).
To further analyze the composition of intestinal microbiota in patients with SP, we selected the top 10 microorganisms ranked by relative abundance in both patients with SP and healthy individuals for comparison. We found that there were significant differences in microbial composition between the two groups of fecal samples at the level of phylum, class, order, family, genus, and species (Fig. 2A–2F). At the phylum level, compared to the healthy population, severe pneumonia patients exhibited a reduced relative abundance of Firmicutes and Bacteroidota, and an increased relative abundance of Proteobacteria and Actinobacteriota in their gut microbiota. At the genus level, the healthy

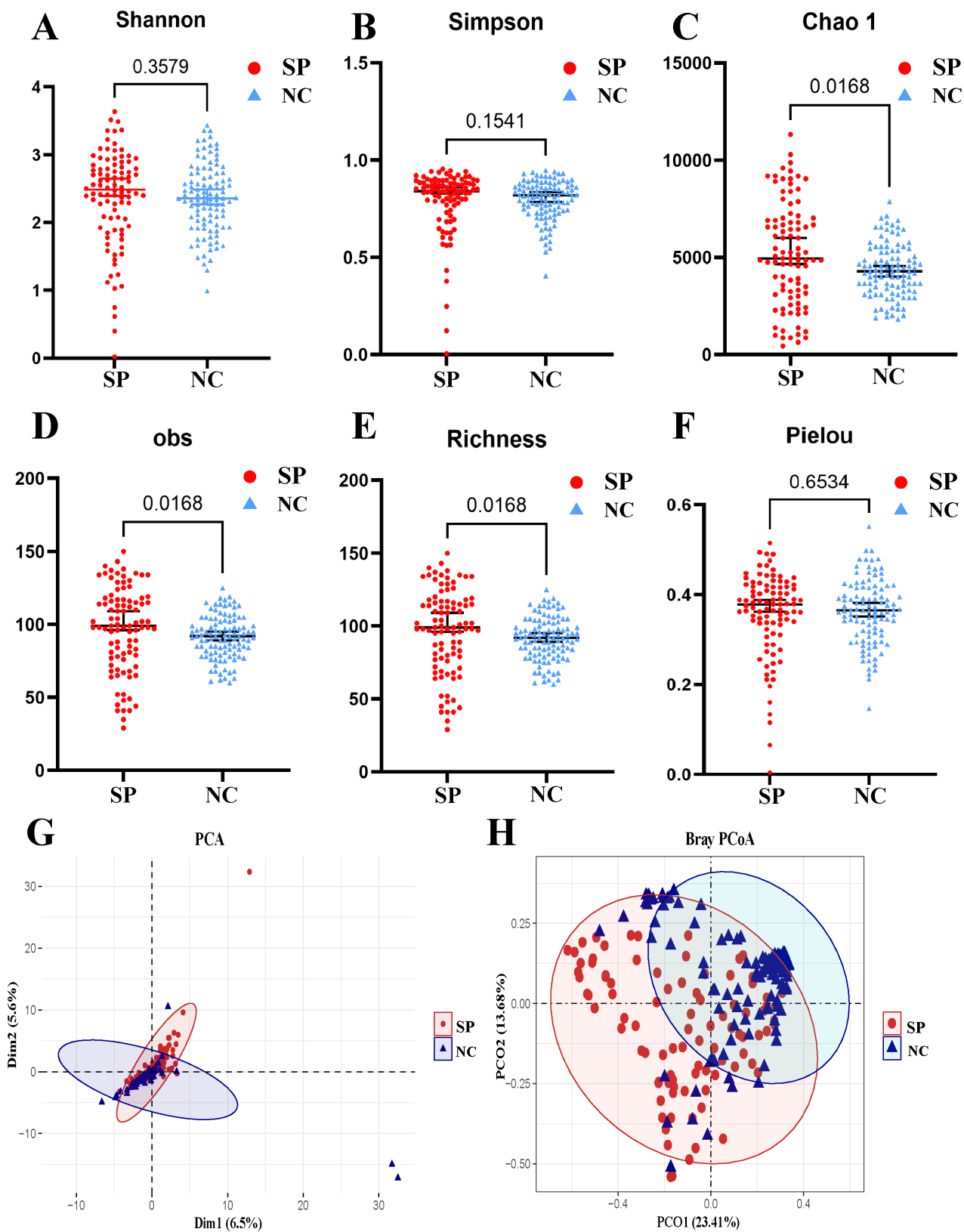


Fig. 1. Microbial diversity analysis of fecal samples. Analysis of α diversity (A–F) and β diversity (G–H) of intestinal microbiota in patients with severe pneumonia (SP) and normal controls (NC).

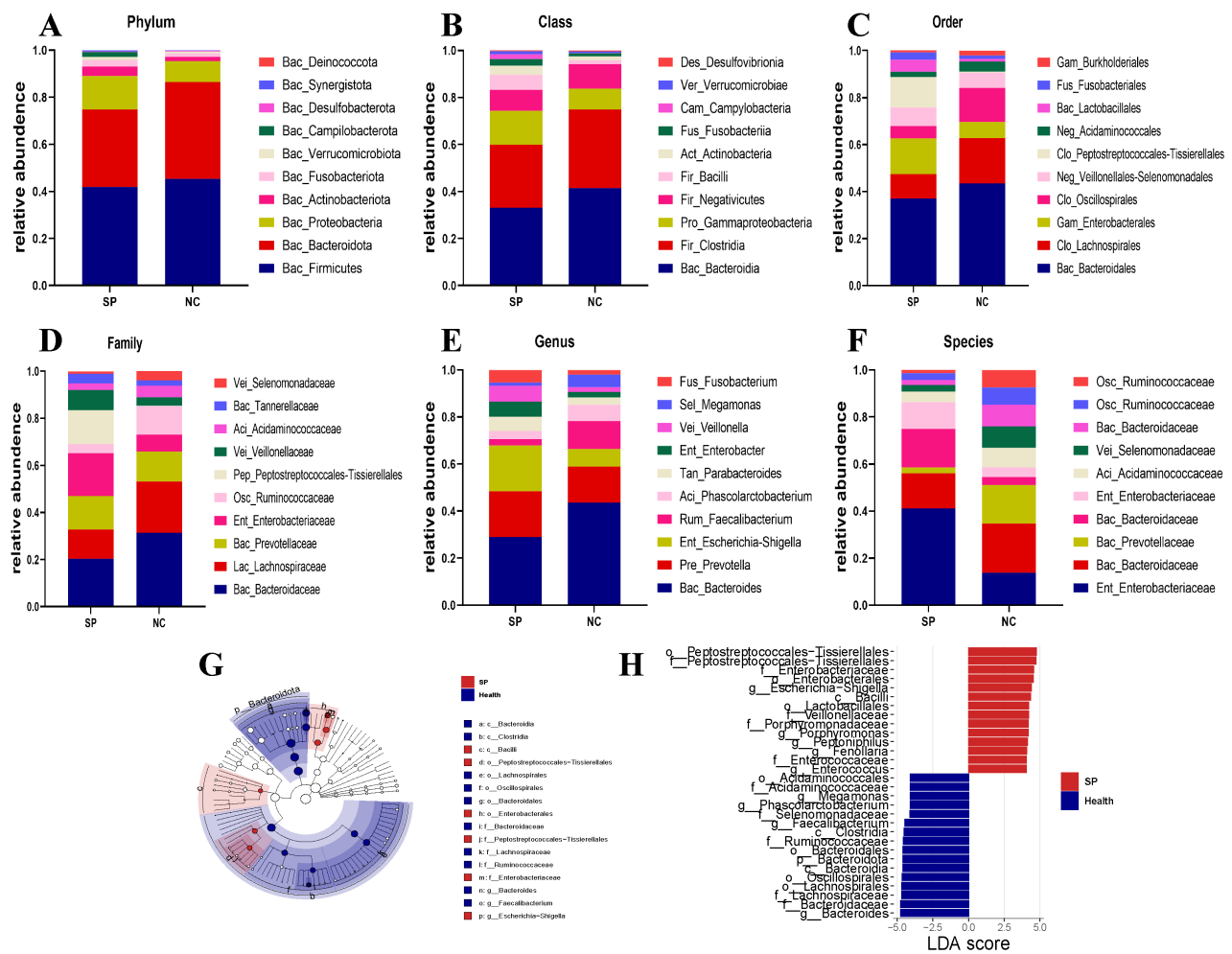


Fig. 2. The microbial composition at different classification levels in fecal samples. Stacked plot of the relative abundances of the top 10 intestinal microorganisms at the phylum (A), class (B), order (C), family (D), genus (E), and species (F) levels. Cladogram (G) and LDA bar plot (H) of microorganisms associated with fecal samples from patients with severe pneumonia and normal controls.

population exhibited a higher abundance of *Bacteroides*, *Faecalibacterium*, *Phascolarctobacterium*, and *Megamonas* in their gut microbiota. At the species level, severe pneumonia patients had a gut microbiota with higher abundances of *Enterobacteriaceae* and *Bacteroidaceae*.

Furthermore, LEfSe analysis revealed significant differences in microbiota abundance at various taxonomic levels between the two groups (Fig. 2G–2H). These results indicate that the gut microbiota of patients with SP appears to differ from that of healthy individuals, providing important insights into the characteristic microbiota of SP patients.

Machine learning methods were used to screen for core microbial markers. To further identify intestinal microbiota biomarkers specific to patients with severe pneumonia, we employed 16S rRNA sequencing to analyze fecal samples and applied multiple machine learning algorithms to identify microbial biomarkers.

The regularization path plot (Fig. 3A) and cross-validation error plot (Fig. 3B) of the LASSO algorithm demonstrate that as the λ value increases, both the model's feature coefficients and the cross-validation tend to converge to zero. In the RF algorithm, as the number of decision trees increases, the model error exhibits a rapid decline followed by stabilization (Fig. 3C). Additionally, we present the importance rankings of the top 15 microbial biomarkers (Fig. 3D). Employing the 5-fold cross-validation methodology, the SVM algorithm demonstrates a decreasing cross-validation error rate alongside increasing accuracy, with the latter stabilizing as the number of microbial biomarkers increases (Fig. 3E–3F). Ultimately, the LASSO algorithm identified 16 genera, the RF algorithm selected 23 genera, and the SVM algorithm screened 101 genera.

Subsequently, we selected the genera that consistently appeared across the microbial communities generated by all three algorithms, ultimately identifying

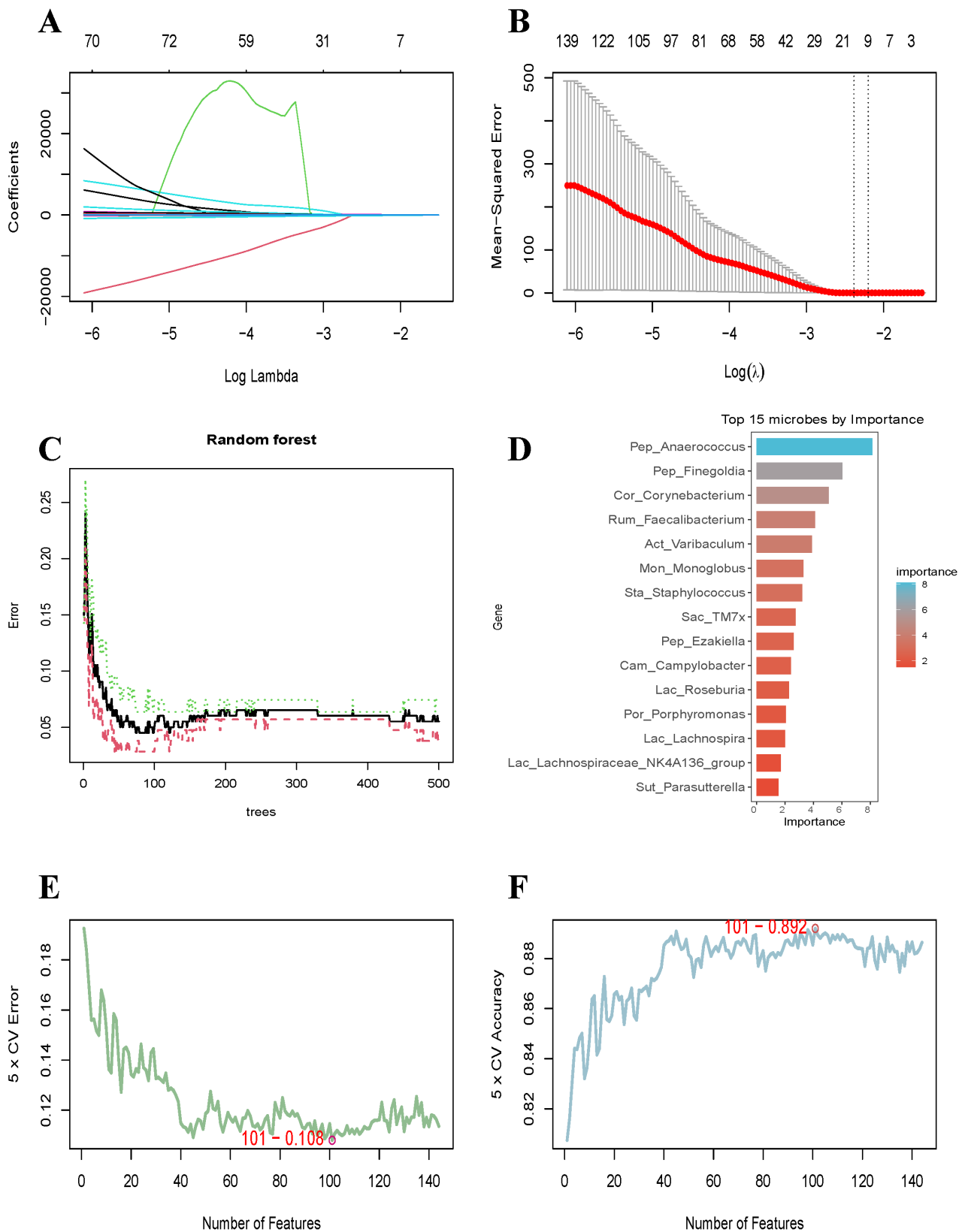


Fig. 3. Machine-learning algorithms screen for signature microbiota. LASSO analysis (A–B), RF analysis (C–D), and SVM analysis (E–F) were used to screen the characteristic microbiota of patients with severe pneumonia at the genus level.

nine microbial genera as characteristic markers for SP patients (Fig. 4A). These nine characteristic microbial markers were *Finegoldia*, *Monoglobus*, *Mogibacterium*, *Peptoniphilus*, *Campylobacter*, *Parasutterella*, *Roseburia*, *Porphyromonas*, and *Faecalibacterium*. The correlation analysis revealed that these nine characteristic microbial markers were not correlated with the blood test parameters (Table SIII and SIV). Comparing the microbial genus abundance between the two groups, we observed that, compared to the healthy

group, the SP group had a reduced relative abundance of *Roseburia*, *Monoglobus*, *Faecalibacterium*, and *Parasutterella*. In contrast, the relative abundance of *Mogibacterium*, *Campylobacter*, *Finegoldia*, *Peptoniphilus*, and *Porphyromonas* was increased (Fig. 4C–4K). We constructed a diagnostic prediction model based on the nine identified microbial genera and generated its corresponding ROC curve (Fig. 4B). The area under the curve (AUC) was approximately 0.969 (95% CI: 0.946–0.992). The model demonstrated 95.24% sen-

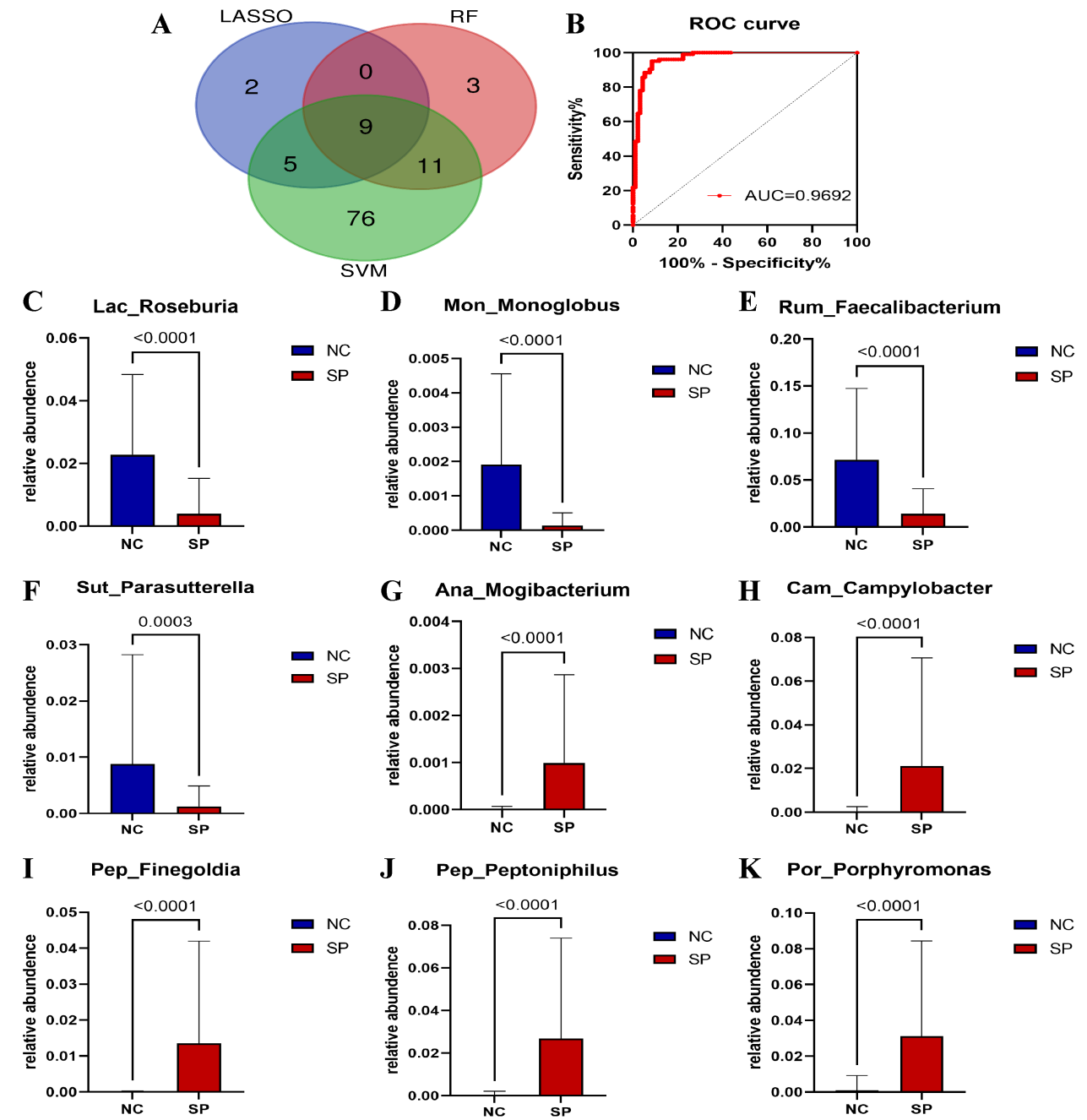


Fig. 4. Evaluation of intestinal characteristic microbial markers in patients with severe pneumonia. The Venn diagram shows the selection of nine microbial markers by three machine learning methods (A). The ROC curve of the combined model of nine microbial markers (B). Changes in the relative abundance of nine microbial markers in patients with severe pneumonia (C–K).

sitivity and 91.49% specificity in diagnosing severe pneumonia. This indicates that the microbiota-based diagnostic prediction model performs well in diagnosing SP patients. Moreover, as characteristic gut microbial genera specific to SP patients, these microbial communities are highly likely to play critical roles in disease progression.

Discussion

Previous studies have demonstrated that interactions between the host and gut microbiota can influence the occurrence and progression of diseases. Specifically, the host's physiological status modulates the homeostasis of the gut microbiota, while alterations in gut microbial communities can indirectly affect the host's disease state through immune responses (Marsland et al. 2015). In this study, we collected clinical data from patients with severe pneumonia to characterize the host's physiological status and further applied machine learning algorithms to analyze the compositional features of the host's gut microbiota. We aimed to seek evidence of the interaction between the intestine and the lungs, and to explore the mechanism of the lung-intestine axis in severe pneumonia.

Our research has revealed that the diversity of the intestinal microbiota in patients with SP has changed. In the alpha diversity analysis, we found that the SP group exhibited significantly higher values for Richness, Chao1, and Obs indices than the normal group. This is a rather unique finding. This is because the body is in an inflammatory state, leading to specific fluctuations in the intestinal microbiota, increased species diversity (especially harmful bacteria), and thus a higher alpha diversity trend compared to the normal group. The β -diversity analysis also revealed differences in species distribution between the two groups.

Using a machine learning algorithm, we screened nine distinct gut microbiota compositions at the genus level for characteristic microbial markers that could identify patients with SP. The general trend in the relative abundance of these nine characteristic microorganisms showed an increase in the relative abundance of harmful bacteria, such as *Finegoldia*, *Mogibacterium*, *Peptoniphilus*, *Campylobacter*, and *Porphyromonas*, while the abundance of beneficial bacteria, such as *Monoglobus*, *Parasutterella*, *Roseburia*, and *Faecalibacterium*, decreased. Previous studies have shown that gut microbiota can play a role in resisting the devel-

opment of disease-related inflammation in the body (Pickard et al. 2017). For example, *Faecalibacterium* is a genus associated with overall health, and its low abundance is linked to the onset and progression of inflammation (Martín et al. 2023). *Porphyromonas* has been found to be associated with several systemic diseases, including cardiovascular disease, Alzheimer's disease, and rheumatoid arthritis. Ju et al. isolated *Parasutterella* from the core intestinal microbiome of mice, and a series of characterization experiments demonstrated its significant role in metabolic processes, offering valuable insights into host health metabolism (Ju et al. 2019). *Campylobacter*, one of the four leading causes of gastroenteritis worldwide, plays a significant role in promoting inflammation (Costa and Iraola 2019). Gou et al. (2023) also found, through animal experiments, that intestinal pathogens, including *Campylobacter* and *Peptoniphilus*, could be depleted after drug intervention, thereby reducing intestinal inflammation and protecting the intestinal barrier. Therefore, we speculate that the microbial marker identified in this study using machine learning algorithms will help alleviate inflammation in severe pneumonia.

Previous studies have focused on various complications of pneumonia, including cardiac involvement. Xu et al. (2020) found that elevated CRP levels, advanced age, underlying comorbidities, and the severity of COVID-19 were the main risk factors for cardiac involvement in patients with COVID-19. Tsuchihashi et al. (2021) analyzed the relationship between clinical characteristics of novel coronavirus infection and severe outcomes and found that advanced age, male sex, and underlying disease were associated with severe outcomes. Similarly, our study collected clinical data for logistic regression and found that age and the total length of hospital stay were significant predictors of recovery and discharge in patients with severe pneumonia.

Among them, age is a significant risk factor for recovery and discharge in patients with severe pneumonia, indicating that older patients are less likely to recover and be discharged. Previous studies have shown that the elderly population has notable differences compared to younger individuals in immune system function, metabolic capacity, and gut microbiota stability (Yu et al. 2024). As individuals age, immune function gradually declines, and the abundance of probiotics in the gut microbiome decreases (Vemuri et al. 2018), which may reduce the body's ability to resist pneumonia. In this study, the total length of hospi-

tal stay was identified as a protective factor for patient improvement and discharge from severe pneumonia. This may be because a more extended hospital stay provides patients with more opportunities to receive comprehensive treatment, including nutritional supplementation, medication adjustments, physical therapy, psychological support, and health education. These interventions help maintain normal metabolism and promote the body's recovery process, including the restoration of gut probiotics. This enables the timely regulation of gut microbiota imbalances, contributing to the patient's recovery and meeting the discharge criteria. Furthermore, the target population of this study comprises patients with severe pneumonia, who require an extended treatment and recovery period. A longer hospital stay may reflect timely intervention and monitoring, offering patients more opportunities for recovery and increasing their chances of meeting the discharge criteria, despite the severity of their condition.

This study investigated the correlation between gut microbiota characteristics and severe pneumonia patients, though several limitations remain. First, the research was conducted at a single center with all samples collected from the same location. All machine learning analyses were performed on a single dataset, which significantly limits the study's generalizability and carries the risk of overfitting. Second, mild cases were excluded because the project primarily focused on critical care patients, leading to insufficient analysis of microbial profiles in these cases. Third, the lack of clinical test data from healthy controls prevented us from identifying differences in clinical factors between patients and healthy individuals.

Although grouping based on "Whether improved and discharged" could provide some evidence for the research, further improvement in the health control group testing is still necessary in subsequent studies.

When establishing a diagnostic prediction model, other risk factors, such as genetic and clinical indicators, were not included in the analysis, and the disease prediction based solely on the nine core microbial markers selected still has some errors. Additionally, changes in gut microbiota are influenced by multiple factors, including antibiotic use, dietary habits, and gastrointestinal disorders.

Future studies should incorporate samples from more centers to refine machine learning results. If possible, Blood samples from healthy populations should be collected for testing, and multi-center cohort stud-

ies should further explore causal relationships between gut microbiota and blood test indicators. Furthermore, precise microecological regulation remains a critical challenge for improving patient outcomes. Future research could explore probiotics, prebiotics, or fecal microbiota transplantation to restore gut balance and enhance immune function, thereby improving clinical outcomes in patients with severe pneumonia.

Conclusion

In summary, we identified a set of gut microbial markers at the genus level that could effectively identify patients with severe pneumonia by using multiple machine learning algorithms. Compared with the normal population, the intestinal microbiota of patients with severe pneumonia has changed. By integrating gut microbiota analysis with patients' clinical profiles, we explored the potential role of the microbiota in the onset of the disease. Our findings will contribute to future research on the gut-lung axis and provide valuable insights for clinical diagnosis and treatment.

Abbreviations

AUC – area under the curve
 CRP – C-reactive protein
 LASSO – Least Absolute Shrinkage and Selection Operator
 LEfSe analysis – LDA Effect Size Analysis
 OTU – operational taxonomic unit
 PCA – principal component analysis
 PCoA – principal co-ordinate analysis
 RF – random forest
 ROC curve – Receiver Operating Characteristic curve
 SP – severe pneumonia
 SVM – support vector machine

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request. Sequencing data associated with this study have been uploaded to the NCBI database (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA1293935?reviewer=mbf6j80s93cgk50pqtbg2kh7f>).

Acknowledgements

The authors would like to express our sincere thanks to all the participants of the People's Hospital of Rongshui and Miao Autonomous County in Guangxi Zhuang Autonomous Region, for the authorization provided by Changhua Wei, Ju Liang, Shudong Wei for the collection of clinical data and samples and the convenience provided to our researchers in this project.

Author contributions

Fangchao Zhong is responsible for data analysis and writing the manuscript. Maosen Huang, Xiaoxia Wei and Lihua Fu collected and organized the data, reviewed and edited the manuscript. Linhai Yan participate in the design and supervision of the full text and provided the main source of funding for the article.

Funding

Guangxi Science and Technology Program (AD25069015), Natural Science Foundation of Guangxi (2023GXNSFDA026032 and 2024GXNSFBA010083), National Natural Science Foundation of China (82160495), Wu Jieping Medical Foundation Research Special Fund (320.6750.2025-02-2).

Conflict of interest

The authors do not report any financial or personal connections with other persons or organizations, which might negatively affect the contents of this publication and/or claim authorship rights to this publication.

Literature

- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghailith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al.** Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019 Aug;37(8):852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- Budden KE, Romani L.** Editorial: Lung microbiome in health and disease. *Front Pharmacol.* 2025 Feb;16:1565849. <https://doi.org/10.3389/fphar.2025.1565849>
- Bulanda E, Wypych TP.** Bypassing the gut-lung axis via microbial metabolites: Implications for chronic respiratory diseases. *Front Microbiol.* 2022 May;13:857418. <https://doi.org/10.3389/fmicb.2022.857418>
- Chen H, Jiang W.** Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Front Microbiol.* 2014 Oct;5:508. <https://doi.org/10.3389/fmicb.2014.00508>
- Costa D, Iraola G.** Pathogenomics of emerging *Campylobacter* species. *Clin Microbiol Rev.* 2019 Jul;32(4):e00072-18. <https://doi.org/10.1128/cmr.00072-18>
- Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE.** Metagenomic analysis of the human distal gut microbiome. *Science.* 2006 Jun;312(5778):1355–1359. <https://doi.org/10.1126/science.1124234>
- Gomes AC, Hoffmann C, Mota JF.** The human gut microbiota: Metabolism and perspective in obesity. *Gut Microbes.* 2018 Jul;9(4):308–325. <https://doi.org/10.1080/19490976.2018.1465157>
- Gou H, Su H, Liu D, Wong CC, Shang H, Fang Y, Zeng X, Chen H, Li Y, Huang Z, et al.** Traditional medicine Pien Tze Huang suppresses colorectal tumorigenesis through restoring gut microbiota and metabolites. *Gastroenterology.* 2023 Dec;165(6):1404–1419. <https://doi.org/10.1053/j.gastro.2023.08.052>
- Gutiérrez-Castrellón P, Gandara-Martí T, Abreu Y Abreu AT, Nieto-Rufino CD, López-Orduña E, Jiménez-Escobar I, Jiménez-Gutiérrez C, López-Velazquez G, Espadaler-Mazo J.** Probiotic improves symptomatic and viral clearance in Covid19 outpatients: A randomized, quadruple-blinded, placebo-controlled trial. *Gut Microbes.* 2022 Jan–Dec;14(1):2018899. <https://doi.org/10.1080/19490976.2021.2018899>
- Hashimoto K.** Emerging role of the host microbiome in neuropsychiatric disorders: Overview and future directions. *Mol Psychiatry.* 2023 Sep;28(9):3625–3637. <https://doi.org/10.1038/s41380-023-02287-6>
- Honarpisheh P, Bryan RM, McCullough LD.** Aging microbiota-gut-brain axis in stroke risk and outcome. *Circ Res.* 2022 Apr;130(8):1112–1144. <https://doi.org/10.1161/circresaha.122.319983>
- Iatcu CO, Steen A, Covasa M.** Gut microbiota and complications of type-2 diabetes. *Nutrients.* 2021 Dec;14(1):166. <https://doi.org/10.3390/nu14010166>
- Ju T, Kong JY, Stothard P, Willing BP.** Defining the role of *Parasutterella*, a previously uncharacterized member of the core gut microbiota. *ISME J.* 2019 Jun;13(6):1520–1534. <https://doi.org/10.1038/s41396-019-0364-5>
- Li Y, Fang M, Li D, Wu P, Wu X, Xu X, Ma H, Li Y, Zhang N.** Association of gut microbiota with critical pneumonia: A two-sample Mendelian randomization study. *Medicine.* 2024 Oct;103(42):e39677. <https://doi.org/10.1097/md.00000000000039677>
- Marsland BJ, Trompette A, Gollwitzer ES.** The gut-lung axis in respiratory disease. *Ann Am Thorac Soc.* 2015 Nov;12(Suppl_2):S150–S156. <https://doi.org/10.1513/annalsats.201503-133aw>
- Martín R, Rios-Covian D, Huillet E, Auger S, Khazaal S, Bermúdez-Humarán LG, Sokol H, Chatel JM, Langella P.** *Faecalibacterium*: A bacterial genus with promising human health applications. *FEMS Microbiol Rev.* 2023 Jul;47(4):fuad039. <https://doi.org/10.1093/femsre/fuad039>
- Nagata N, Takeuchi T, Masuoka H, Aoki R, Ishikane M, Iwamoto N, Sugiyama M, Suda W, Nakanishi Y, Terada-Hirashima J, et al.** Human gut microbiota and its metabolites impact immune responses in COVID-19 and its complications. *Gastroenterology.* 2023 Feb;164(2):272–288. <https://doi.org/10.1053/j.gastro.2022.09.024>
- Pickard JM, Zeng MY, Caruso R, Núñez G.** Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol Rev.* 2017 Sep;279(1):70–89. <https://doi.org/10.1111/imr.12567>
- Qiu P, Ishimoto T, Fu L, Zhang J, Zhang Z, Liu Y.** The gut microbiota in inflammatory bowel disease. *Front Cell Infect Microbiol.* 2022 Feb;12:733992. <https://doi.org/10.3389/fcimb.2022.733992>
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.** The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013 Jan;41(D1):D590–D596. <https://doi.org/10.1093/nar/gks1219>
- R Core Team.** A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2024 [cited 2025 May 05]. Available from <https://www.r-project.org>
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F.** VSEARCH: A versatile open source tool for metagenomics. *PeerJ.* 2016 Oct;4:e2584. <https://doi.org/10.7717/peerj.2584>
- Silverio R, Gonçalves DC, Andrade MF, Seelaender M.** Coronavirus disease 2019 (COVID-19) and nutritional status: The missing link? *Adv Nutr.* 2021 Jun;12(3):682–692. <https://doi.org/10.1093/advances/nmaa125>
- Tian C, Li X, Yuan X, Li Z.** Link between respiratory microbiota and asthma: An emerging therapeutic approach. *Am J Transl Res.* 2024 Nov;16(11):6289–6302. <https://doi.org/10.62347/osnz1705>
- Torres A, Cilloniz C, Niederman MS, Menéndez R, Chalmers JD, Wunderink RG, van der Poll T.** Pneumonia. *Nat Rev Dis Primers.* 2021 Apr 8;7(1):25. <https://doi.org/10.1038/s41572-021-00259-0>
- Tsuchihashi Y, Arima Y, Takahashi T, Kanou K, Kobayashi Y, Sunagawa T, Suzuki M.** Clinical characteristics and risk factors for severe outcomes of novel coronavirus infection, January–March 2020, Japan. *J Epidemiol.* 2021 Aug;31(8):487–494. <https://doi.org/10.2188/jea.je20200519>
- Vemuri R, Gundamaraju R, Shastri MD, Shukla SD, Kalpurath K, Ball M, Tristram S, Shankar EM, Ahuja K, Eri R.** Gut microbial changes, interactions, and their implications on human lifecycle: An ageing perspective. *Biomed Res Int.* 2018 Feb;2018:4178607. <https://doi.org/10.1155/2018/4178607>

- Wang L, Cai Y, Garssen J, Henricks PAJ, Folkerts G, Braber S. The bidirectional gut-lung axis in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2023 May;207(9):1145–1160. <https://doi.org/10.1164/rccm.202206-1066tr>
- Wong CC, Yu J. Gut microbiota in colorectal cancer development and therapy. *Nat Rev Clin Oncol*. 2023 Jul;20(7):429–452. <https://doi.org/10.1038/s41571-023-00766-x>
- Xu H, Hou K, Xu R, Li Z, Fu H, Wen L, Xie L, Liu H, Selvanayagam JB, Zhang N, et al. Clinical characteristics and risk factors of cardiac involvement in COVID-19. *J Am Heart Assoc*. 2020 Sep;9(18):e016807. <https://doi.org/10.1161/jaha.120.016807>
- Yang W, Cong Y. Gut microbiota-derived metabolites in the regulation of host immune responses and immune-related inflammatory diseases. *Cell Mol Immunol*. 2021 Apr;18(4):866–877. <https://doi.org/10.1038/s41423-021-00661-4>
- Yu L, Pan J, Guo M, Duan H, Zhang H, Narbad A, Zhai Q, Tian F, Chen W. Gut microbiota and anti-aging: Focusing on spermidine. *Crit Rev Food Sci Nutr*. 2024;64(28):10419–10437. <https://doi.org/10.1080/10408398.2023.2224867>
- Zhao L, Zhang X, Zhou Y, Fu K, Lau HC, Chun TW, Cheung AH, Coker OO, Wei H, Wu WK, et al. *Parvimonas micra* promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients. *Oncogene*. 2022 Sep;41(36):4200–4210. <https://doi.org/10.1038/s41388-022-02395-7>

Supplementary materials are available on the journal's website.