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.

# A Biologically Interpretable Multimodal Model for Prognostic Risk Stratification and Glucocorticoid-Associated Outcome Heterogeneity in Acute Respiratory Distress Syndrome (ARDS)

**Zhenju Song**

song.zhenju@zs-hospital.sh.cn

Zhongshan Hospital, Fudan University https://orcid.org/0000-0001-9294-8824

**Mengna Lin**

Zhongshan Hospital, Fudan University

**Su Lu**

Zhongshan Hospital, Fudan University

**Yan Zeng**

Shanghai United Imaging Intelligence Co., Ltd.

**Na Zheng**

Shanghai Institute of Infectious Disease and Biosecurity, Fudan University

**Yumei Chen**

Zhongshan Hospital, Fudan University

**Chen Chen**

Zhongshan Hospital, Fudan University

**Xiaojing Zou**

Wuhan Union Hospital

**Feixiang Xu**

Zhongshan Hospital, Fudan University

**Zhiwei Chen**

Zhongshan Hospital, Fudan University

**Simeng Wu**

Zhongshan Hospital, Fudan University

**Chengjin Gao**

Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine

**Xueling Wu**

Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine

**Zhaocai Zhang**

Second Affiliated Hospital，Zhejiang University School of Medicine

**Qi Liu**

The First Affiliated Hospital of Zhengzhou University https://orcid.org/0000-0002-9868-3705

**Yao Shen**

Shanghai Pudong Hospital

**Yiyan Lin**

Guangdong Provincial People's Hospital

**Yiyu Deng**

Children's Hospital Boston, Harvard Medical School

**Ran Ji**

Second Affiliated Hospital Zhejiang University School of Medicine

**Jianfeng Song**

Minhang Hospital,Fudan University

**Lu Wang**

Renmin Hospital of Wuhan University

**Chengchao Peng**

Union Hospital, Tongji Medical College, Huazhong University of Science and Technology

**Xue Tian**

Union Hospital, Tongji Medical College, Huazhong University of Science and Technology

**Ruilan Wang**

Department of Critical Care Medicine, Shanghai General Hospital, Shanghai Jiaotong University, School of Medicine

**Lei Chen**

Shanghai United Imaging Intelligence Co., Ltd.

**Liying Zhan**

Renmin Hospital of Wuhan University

---

**Article**

**Keywords:**

**Posted Date:** November 1st, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11277228/v1

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

**Additional Declarations:** There is **NO** Competing Interest.

---

# Abstract

The clinical and biological heterogeneity of acute respiratory distress syndrome (ARDS) complicates early prognostic assessment and may contribute to variability in treatment-associated outcomes. In this multicentre study of 2,868 patients from 10 centres, we developed and externally validated a multimodal risk-stratification framework integrating clinical variables, chest CT imaging, and serum proteomic data. Serum proteomic profiles were available in 879 patients and exploratory bronchoalveolar lavage fluid (BALF) single-cell RNA sequencing was performed in 26 patients to characterize pulmonary microenvironment of model-defined risk states. The clinical-imaging model achieved AUROCs of 0.831, 0.805, and 0.781 for 30-day mortality prediction in the training, internal test, and external validation cohorts, respectively, while proteomic integration improved discrimination to 0.892, 0.871, and 0.859 in the proteome-matched subset. Higher predicted risk was associated with impaired early liberation from mechanical ventilation and increased non-invasive ventilation failure. Serum proteomic analyses revealed distinct biological profiles of high-risk group characterized by tissue-injury and inflammatory signatures, such as KRT19, CHIT1, IGFBP2, and IL-6, with enrichment of inflammatory pathways (HMGB1 and TREM1 signalling), metabolic and nuclear receptor pathways (FXR/RXR and PPAR signalling), extracellular matrix remodelling (FAK signalling and collagen degradation), endothelial dysfunction, and coagulation pathways. BALF single-cell RNA further identified neutrophil-enriched pulmonary immune responses in high-risk patients. Glucocorticoid-associated 30-day mortality varied across model-defined risk strata (*P* for interaction < 0.001 and 0.033 for the clinical-imaging and multimodal models, respectively), with lower mortality associated with glucocorticoid initiation among high-risk patients. These findings support multimodal risk stratification as a framework for identifying clinically and biologically distinct ARDS risk states and generating hypotheses regarding treatment-associated heterogeneity.


# Introduction

Acute respiratory distress syndrome (ARDS) is a life-threatening clinical condition characterized by acute-onset hypoxemia, bilateral pulmonary infiltrates, and non-cardiogenic pulmonary edema[1, 2]. Despite advances in supportive care, ARDS continues to carry a high mortality rate, ranging from 30% to 40%, particularly in patients requiring invasive mechanical ventilation[1]. The clinical heterogeneity and dynamic progression of ARDS present substantial challenges to early risk stratification and individualized treatment strategies[3]. Early prognostic assessment is essential for clinical triage, enabling timely identification of patients at high risk of death who may benefit from closer monitoring and more intensive management[4]. Traditional prognostic tools, such as the $PaO_2/FiO_2$ ratio, SOFA score, and APACHE II score, offer limited predictive performance due to their reliance on general physiological variables and inability to reflect underlying molecular or pulmonary structural alterations[5, 6, 7, 8]. Therefore, there is a pressing need for more precise, biologically informed models capable of early outcome prediction.

Radiomics and advanced image-analysis techniques enable quantitative characterization of the spatial and structural patterns of lung injury on chest computed tomography (CT)[9]. CT therefore provides complementary macroscopic information beyond conventional clinical variables and has shown promise for early risk stratification in other critical illnesses, including sepsis[10, 11]. Artificial intelligence (AI) further extends this potential by enabling automated extraction of high-dimensional imaging features, while radiomics quantifies lung morphology, texture, and heterogeneity[12]. However, quantitative CT has been incorporated into relatively few ARDS prognostic models, particularly for predicting prolonged invasive mechanical ventilation (IMV). Moreover, the generalizability of AI-

based models across heterogeneous patient populations and clinical settings remains an important concern, highlighting the need for robust external validation[13].

Advances in high-throughput omics provide an opportunity to complement conventional clinical and imaging assessment with molecular characterization of ARDS[2, 14]. Blood proteomics captures systemic inflammatory and tissue-injury processes, and bronchoalveolar lavage fluid (BALF) single-cell RNA sequencing (scRNA-seq) resolves pulmonary cellular states and immune microenvironmental changes [15]. However, most existing studies remain limited to single-modality analyses[10, 13, 16], which fail to fully capture the complex structural, molecular, and cellular heterogeneity of ARDS. In addition, the “black-box” nature of many AI models limits their interpretability and clinical translatability[17]. A biologically interpretable multimodal framework may therefore enable more precise risk stratification and identification of patients with differential treatment responses.

We hypothesized that integrating clinical, CT imaging, and blood proteomic data could identify robust, biologically meaningful risk states in ARDS. We therefore developed and externally validated a multimodal model for 30-day mortality, assessed its clinical relevance using respiratory-support outcomes, and characterized associated biological features using serum proteomics and exploratory BALF single-cell RNA sequencing. We further explored whether model-defined risk modified outcomes associated with glucocorticoid treatment.

## Results

# Study flow and participant characteristics

Figure 1A shows the study flow and cohort assembly. 3,451 potentially eligible patients were screened across 10 hospitals. After application of the criteria, 2,868 patients were included in the final analysis, and their baseline clinical characteristics are summarized in Table 1. The study included 1,715 patients in the training cohort (mean age, 68.6 ± 15.1 years), 326 in the internal test cohort (mean age, 68.1 ± 15.4 years), and 827 in the external validation cohort (mean age, 64.2 ± 15.4 years). 30-day mortality was 38.0%, 35.3%, and 33.9% in training, internal test, and external validation cohort, respectively.

Table 1
Baseline characteristics of the study population

| **Sex (Female, %)** | **Overall (n = 2868)** | **Training cohort (n = 1715)** | **Internal test cohort (n = 326)** | **External validation (n = 827)** | **P** |
|---|---|---|---|---|---|
| | **881 (30.7)** | **507 (29.6)** | **109 (33.4)** | **265 (32.0)** | **0.231** |
| Age (mean (SD)) | 67.2 (15.3) | 68.6 (15.1) | 68.1 (15.4) | 64.2 (15.4) | < 0.001 |
| SOFA | 7 [4, 9] | 8 [6, 10] | 6 [4, 9] | 6 [4, 9] | < 0.001 |
| P/F ratio | 166 [111, 233] | 167 [112, 230] | 158 [100, 213] | 171 [112, 245] | 0.036 |
| Comorbidity (n, %) | | | | | |
| Hypertension | 1486 (51.8) | 918 (53.5) | 156 (47.9) | 412 (49.8) | 0.067 |
| Coronary heart disease | 404 (14.1) | 222 (12.9) | 56 (17.2) | 126 (15.2) | 0.069 |
| Chronic heart failure | 456 (15.9) | 305 (17.8) | 31 (9.5) | 120 (14.5) | < 0.001 |
| COPD | 190 (6.6) | 135 (7.9) | 19 (5.8) | 36 (4.4) | 0.003 |
| Asthma | 48 (1.7) | 30 (1.7) | 7 (2.1) | 11 (1.3) | 0.549 |
| Interstitial Lung Disease | 105 (3.7) | 46 (2.7) | 35 (10.7) | 24 (2.9) | < 0.001 |
| Diabetes | 831 (29.0) | 500 (29.2) | 100 (30.7) | 231 (27.9) | 0.628 |
| Chronic kidney disease | 406 (14.2) | 276 (16.1) | 34 (10.4) | 96 (11.6) | 0.001 |
| Liver cirrhosis | 90 (3.1) | 55 (3.2) | 9 (2.8) | 26 (3.1) | 0.962 |
| Hematologic malignancy | 109 (3.8) | 68 (4.0) | 10 (3.1) | 31 (3.7) | 0.789 |
| Cerebrovascular disease | 574 (20.0) | 364 (21.2) | 60 (18.4) | 150 (18.1) | 0.146 |
| Autoimmune disease | 156 (5.4) | 97 (5.7) | 24 (7.4) | 35 (4.2) | 0.087 |
| Immunosuppressant use | 218 (7.6) | 150 (8.7) | 33 (10.1) | 35 (4.2) | < 0.001 |
| Solid organ transplantation | 91 (3.2) | 52 (3.0) | 13 (4.0) | 26 (3.1) | 0.637 |
| Solid tumor | 492 (17.2) | 293 (17.1) | 50 (15.3) | 149 (18.0) | 0.562 |
| Chronic lung disease | 316 (11.0) | 190 (11.1) | 56 (17.2) | 70 (8.5) | < 0.001 |
| Direct ARDS | 2223 (77.5) | 1439 (84.0) | 283 (86.8) | 501 (60.6) | < 0.001 |
| Day30 mortality | 1047 (36.5) | 652 (38.0) | 115 (35.3) | 280 (33.9) | 0.110 |
| Day60 mortality | 1322 (46.1) | 845 (49.3) | 149 (45.7) | 328 (39.7) | < 0.001 |

Abbreviations: SOFA, Sequential Organ Failure Assessment; P/F ratio, $PaO_2/FiO_2$ ratio; COPD: Chronic Obstructive Pulmonary Disease; WBC, White Blood Cell Count; CRP, C-reactive Protein; PCT, Procalcitonin; RR, Respiratory Rate; Temp, Body Temperature; HR, Heart Rate; Lac, Lactate; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; cTnT: cardiac Troponin T; Cr, Creatinine; BUN, Blood Urea; STB, Serum Total Bilirubin; AST, Aspartate aminotransferase; ALT: Alanine Aminotransferase; Fg, Fibrinogen; PT, Prothrombin Time; APTT, Activated Partial Thromboplastin Time; PLT, Platelet Count; INR, International Normalized Ratio; Alb, Albumin; HCT, Hematocrit; BE, Base Excess; SD, Standard Deviation.

| Sex (Female, %) | Overall (n = 2868) | Training cohort (n = 1715) | Internal test cohort (n = 326) | External validation (n = 827) | P |
|---|---|---|---|---|---|
| | **881 (30.7)** | **507 (29.6)** | **109 (33.4)** | **265 (32.0)** | **0.231** |
| Day90 mortality | 1418 (49.4) | 897 (52.3) | 168 (51.5) | 353 (42.7) | < 0.001 |
| Day180 mortality | 1517 (52.9) | 961 (56.0) | 176 (54.0) | 380 (45.9) | < 0.001 |
| Vital signs and lab tests | | | | | |
| WBC ($\times 10^9$/L) | 10.3 [7.0, 14.8] | 10.2 [7.0, 14.7] | 10.3 [7.0, 14.4] | 10.7 [7.2, 14.9] | 0.321 |
| CRP (mg/L) | 90.0 [40.4, 157.0] | 90.0 [37.0, 158.8] | 90.0 [43.2, 166.4] | 90.0 [44.7, 151.1] | 0.241 |
| PCT (ng/mL) | 0.9 [0.2, 4.9] | 0.7 [0.2, 3.9] | 0.9 [0.2, 4.9] | 1.6 [0.3, 7.0] | < 0.001 |
| RR (breaths/min) | 22 [19, 27] | 20 [19, 25] | 22 [20, 29] | 22 [18, 38] | < 0.001 |
| Temp (°C) | 37.0 [36.6, 37.7] | 37.0 [36.6, 37.8] | 37.1 [36.7, 38.0] | 36.8 [36.5, 37.5] | < 0.001 |
| HR (beats/min) | 96 [80, 110] | 93 [80, 110] | 100 [83, 115] | 98 [83, 113] | < 0.001 |
| Lac (mmol/L) | 1.6 [1.2, 2.4] | 1.6 [1.2, 2.4] | 1.8 [1.3, 2.7] | 1.6 [1.1, 2.4] | < 0.001 |
| SBP (mmHg) | 123 [107, 140] | 124 [110, 140] | 123 [104, 140] | 120 [101, 139] | < 0.001 |
| DBP (mmHg) | 68 [58, 79] | 70 [60, 80] | 70 [60, 80] | 62 [42, 78] | < 0.001 |
| cTnT (ng/L) | 0.0 [0.0, 0.1] | 0.0 [0.0, 0.1] | 0.0 [0.0, 0.1] | 0.0 [0.0, 0.1] | < 0.001 |
| Cr (μmol/L) | 85.7 [61.0, 145.0] | 86.0 [62.0, 144.0] | 84.5 [59.0, 135.5] | 86.0 [58.0, 159.6] | 0.662 |
| BUN (mmol/L) | 10.2 [6.7, 17.6] | 10.2 [6.6, 16.9] | 9.8 [6.3, 16.9] | 10.5 [7.1, 18.3] | 0.083 |
| Urine volume (mL) | 1450 [879, 2050] | 1480 [950, 2000] | 1450 [900, 2030] | 1450 [750, 2214] | 0.961 |
| STB (μmol/L) | 13.9 [8.9, 22.9] | 12.9 [8.4, 21.6] | 13.9 [9.0, 20.8] | 16.2 [10.5, 27.4] | < 0.001 |
| AST (U/L) | 36.0 [23.0, 64.0] | 36.0 [23.0, 61.0] | 38.0 [25.0, 68.8] | 36.0 [22.0, 68.5] | 0.069 |
| ALT (U/L) | 27.0 [16.0, 50.0] | 28.0 [17.0, 50.0] | 27.5 [18.0, 52.0] | 25.0 [14.0, 50.6] | 0.027 |

Abbreviations: SOFA, Sequential Organ Failure Assessment; P/F ratio, $PaO_2/FiO_2$ ratio; COPD: Chronic Obstructive Pulmonary Disease; WBC, White Blood Cell Count; CRP, C-reactive Protein; PCT, Procalcitonin; RR, Respiratory Rate; Temp, Body Temperature; HR, Heart Rate; Lac, Lactate; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; cTnT: cardiac Troponin T; Cr, Creatinine; BUN, Blood Urea; STB, Serum Total Bilirubin; AST, Aspartate aminotransferase; ALT: Alanine Aminotransferase; Fg, Fibrinogen; PT, Prothrombin Time; APTT, Activated Partial Thromboplastin Time; PLT, Platelet Count; INR, International Normalized Ratio; Alb, Albumin; HCT, Hematocrit; BE, Base Excess; SD, Standard Deviation.

| Sex (Female, %) | Overall (n = 2868) | Training cohort (n = 1715) | Internal test cohort (n = 326) | External validation (n = 827) | P |
|---|---|---|---|---|---|
| | **881 (30.7)** | **507 (29.6)** | **109 (33.4)** | **265 (32.0)** | **0.231** |
| Fg (mg/dL) | 449 [312, 606] | 450 [310, 610] | 453 [332, 585] | 442 [309, 600] | 0.624 |
| PT (s) | 13.9 [12.7, 15.6] | 13.8 [12.6, 15.5] | 13.8 [12.6, 15.4] | 14.3 [12.8, 15.9] | 0.017 |
| APTT (s) | 33.0 [28.8, 39.7] | 31.7 [28.0, 37.3] | 33.6 [29.5, 40.8] | 36.1 [30.6, 43.5] | < 0.001 |
| D-dimer (mg/L) | 4.1 [1.9, 9.1] | 3.9 [1.9, 8.9] | 2.7 [1.1, 7.4] | 4.9 [2.3, 10.0] | < 0.001 |
| PLT ($\times 10^9$/L) | 164 [104, 236] | 168 [106, 233] | 166 [119, 240] | 155 [94, 237] | 0.035 |
| INR | 1.2 [1.1, 1.3] | 1.2 [1.1, 1.3] | 1.2 [1.1, 1.3] | 1.2 [1.1, 1.3] | 0.078 |
| Alb (g/L) | 31.0 [28.0, 34.0] | 31.0 [28.0, 34.6] | 30.0 [27.0, 34.0] | 30.9 [28.0, 33.7] | < 0.001 |
| HCT (%) | 31.2 [25.2, 36.7] | 32.0 [25.9, 37.2] | 32.8 [27.7, 38.3] | 28.4 [23.5, 34.2] | < 0.001 |
| pH | 7.4 [7.3, 7.5] | 7.4 [7.3, 7.5] | 7.4 [7.4, 7.4] | 7.4 [7.3, 7.5] | 0.892 |
| BE (mmol/L) | -0.6 [-4.0, 3.1] | -0.7 [-4.1, 3.1] | 0.1 [-3.2, 3.0] | -0.5 [-4.1, 3.0] | 0.146 |

Abbreviations: SOFA, Sequential Organ Failure Assessment; P/F ratio, $PaO_2/FiO_2$ ratio; COPD: Chronic Obstructive Pulmonary Disease; WBC, White Blood Cell Count; CRP, C-reactive Protein; PCT, Procalcitonin; RR, Respiratory Rate; Temp, Body Temperature; HR, Heart Rate; Lac, Lactate; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; cTnT: cardiac Troponin T; Cr, Creatinine; BUN, Blood Urea; STB, Serum Total Bilirubin; AST, Aspartate aminotransferase; ALT: Alanine Aminotransferase; Fg, Fibrinogen; PT, Prothrombin Time; APTT, Activated Partial Thromboplastin Time; PLT, Platelet Count; INR, International Normalized Ratio; Alb, Albumin; HCT, Hematocrit; BE, Base Excess; SD, Standard Deviation.

## Prognosis model establishment and performance

Integration of clinical and imaging information improved model discrimination, yielding AUROCs of 0.831 (95% CI, 0.812–0.851), 0.805 (0.756–0.854), and 0.781 (0.748–0.814) in the training, internal test, and external validation cohorts, respectively. The clinical-imaging model outperformed the clinical, radiomics, and deep learning CT models across cohorts (Figs. 2A-C; all bootstrap $P < 0.001$) and showed balanced classification performance across sensitivity, specificity, accuracy, positive predictive value, and negative predictive value (**Table S1**). The model also demonstrated favorable calibration, with corresponding Brier scores of 0.161, 0.193, and 0.178 (**Figure S1**; **Table S2**).

Among patients with matched proteomic data, the proteomic model achieved AUROCs of 0.847 (95% CI, 0.813–0.881), 0.829 (0.769–0.890), and 0.803 (0.725–0.881) in the training, internal test, and external validation cohorts, respectively. The corresponding AUROCs of the clinical-imaging model were 0.858 (0.826–0.890), 0.821 (0.760–

0.884), and 0.814 (0.743–0.885). Integration of proteomic information with clinical and imaging features further improved discrimination, yielding AUROCs of 0.892 (0.865–0.920), 0.871 (0.818–0.923), and 0.859 (0.806–0.922), respectively (Figs. 2D-F**)**. Compared with the clinical-imaging model, the multimodal model increased AUROC by 0.034 (95% CI, 0.017–0.052; bootstrap $P < 0.001$), 0.050 (0.015–0.088; $P = 0.008$), and 0.045 (0.012–0.089; $P = 0.030$) in the training, internal test, and external validation cohorts, respectively (**Figures S2-3**).

The improvement in discrimination was accompanied by lower prediction error, with Brier scores of 0.126, 0.147, and 0.147 across the three cohorts, respectively (**Table S3**). Threshold-dependent classification performance was generally improved with multimodal integration, particularly for specificity, PPV, and overall accuracy, although changes in sensitivity and NPV varied across cohorts (**Table S4**). Decision curve analysis further demonstrated greater net benefit for the multimodal model across clinically relevant threshold probabilities (**Figure S4**). These findings support a consistent incremental contribution of proteomic information beyond clinical and CT-derived features, including during external validation.

In addition, the multimodal model showed sustained discrimination across longer follow-up horizons, with AUCs of 0.871, 0.857, and 0.847 for 60-day mortality; 0.864, 0.851, and 0.844 for 90-day mortality; and 0.851, 0.841, and 0.834 for 180-day mortality in the training, internal test, and external validation cohorts, respectively (Fig. 2G).

## Prognostic performance across clinically relevant subgroups

Model discrimination was evaluated across subgroups defined by age, ARDS etiology, and the presence or absence of chronic pulmonary disease (Fig. 3A). Analyses used the available population for each model, with proteomic and multimodal analyses restricted to patients with matched proteomic data. In the external validation cohort, multimodal-model AUROCs ranged from 0.83 to 0.94 across these subgroups. Estimates were less precise in some smaller subgroups, including patients with chronic pulmonary disease (AUROC, 0.86; 95% CI, 0.74–0.99; n = 25).

## Model interpretation and key contributors

SHAP analysis identified age, $PaO_2/FiO_2$ ratio, eosinophil percentage, solid tumour, lactate, urine output, and blood urea nitrogen as the leading clinical contributors (Figs. 3B-D). Within the radiomic signature, first-order skewness carried the largest negative weight, whereas kurtosis, and GLCM sum entropy captured attenuation heterogeneity, spatial disorganization, and extent of lung involvement; proteomic importance was dominated by KRT19, followed by HEXIM1, CBLIF, TFRC, HLA-E, CKMT1A/CKMT1B, CHIT1, NFATC1, IGFBP2, and IL-6. Together, these findings indicate that the multimodal model integrates complementary physiological, structural, and molecular dimensions of ARDS risk.

## Comparison with physician and ChatGPT mortality prediction

A total of 20 emergency and critical care physicians from five medical centres independently evaluated patients in the internal test cohort (**Figure S5**). Readers had a median of 8 years of clinical experience (interquartile range, 5–14 years). The clinical-imaging model achieved an AUROC of 0.805 (0.756–0.854), compared with a mean AUROC of 0.587 across physician readers; the difference was significant in MRMC analysis ($P < 0.001$). ChatGPT achieved an AUROC of 0.578 (0.514–0.642), which was also lower than that of the clinical-imaging model (paired bootstrap $P < 0.001$). Individual reader performance is shown in **Figure S6**.

# Prediction of early successful extubation and non-invasive mechanical ventilation (NIMV) failure

Among patients with matched proteomic data, the clinical-imaging and multimodal models showed good discrimination for respiratory-support outcomes. For early successful extubation, the clinical-imaging and multimodal models achieved AUCs of 0.836 and 0.812, respectively (Fig. 4A). For NIMV failure, the corresponding AUCs were 0.819 and 0.853 (Fig. 4B). Higher model-predicted mortality risk was consistently associated with lower odds of early successful weaning and higher odds of NIMV failure for both models (Figs. 4C-F; all overall $P < 0.001$). No significant non-linear associations were observed.

## Model-derived risk groups showed heterogeneous associations with glucocorticoid treatment

Model-derived risk groups showed heterogeneous associations with glucocorticoid treatment
Patients were stratified into low-, intermediate-, and high-risk groups according to tertiles of predicted probability, using cutoffs of 0.220 and 0.492 for the clinical-imaging model and 0.161 and 0.491 for the multimodal model. Both models showed clear separation of 90-day survival across the training, internal test, and external validation cohorts (all log-rank $P < 0.001$; Figs. 5A-F). After excluding patients with systemic glucocorticoid exposure before enrollment, eligible patients from all three cohorts were pooled for the treatment-associated outcome analysis. Time-varying glucocorticoid initiation showed heterogeneous associations with 30-day mortality across clinical-imaging risk groups ($P$ for interaction < 0.001), with HRs of 0.70 (95% CI, 0.60–0.82), 1.15 (0.90–1.46), and 1.22 (0.82–1.81) in the high-, intermediate-, and low-risk groups, respectively. A similar gradient was observed for the multimodal model ($P$ for interaction = 0.033) (Fig. 5G). Continuous-risk analyses were consistent with these findings, with lower glucocorticoid-associated HRs at higher predicted risk for both models (both $P$ for interaction < 0.001; **Figure S7**).

## Serum proteome provides biological characterization of model-defined risk states

Differential proteomic analyses were performed across risk strata defined by the clinical-imaging and multimodal models. Notably, substantial overlap was observed between the proteomic alterations associated with the two risk-stratification approaches across pairwise risk-group comparisons (Figs. 6A-D), indicating that clinical and CT information alone captured a considerable component of the molecular gradient identified after proteomic integration. Proteomic integration additionally identified model-specific alterations, suggesting further molecular refinement. Twenty-five proteins were consistently represented across the overlapping differential-protein sets (Fig. 6E), with many showing progressive changes across increasing risk strata. The shared signature encompassed inflammatory and immune mediators (IL-6, IL-17C, IL-10, IL1RL1, and CCL7), epithelial injury and tissue-remodelling proteins (ITGB6, KRT19, GDF15, PDGFA, PTN, and MSTN), and markers of cardiovascular stress (NPPB and NT-proBNP).

Pathway analysis showed broadly concordant biological programs across clinical-imaging- and multimodal-defined risk strata (Fig. 6F). Increasing risk was associated with predicted activation of HMGB1, IL-6, NOD1/2, S100, TREM1, Toll-like receptor, and cytokine-storm-related signalling, together with attenuation of PPAR and LXR/RXR

pathways. Alterations in extracellular-matrix remodelling, focal adhesion, endothelial dysfunction, coagulation, and platelet-related pathways were also observed.

# BALF single-cell transcriptome reveals a neutrophil-dominant and stress-associated pulmonary microenvironment in high-risk ARDS

BALF samples were prospectively obtained from 26 patients with ARDS stratified by the multimodal model, including 13 high-risk, 4 intermediate-risk, and 9 low-risk patients. After quality control, 153,019 cells were retained for dimensionality reduction and clustering, yielding 11 major cell populations based on canonical markers (Fig. 7A; **Figure S8**). High-risk patients showed broad enrichment of neutrophil states, including activated, interferon-responsive, inflammatory, metabolic, and stress-associated populations, whereas multiple macrophage and T/NK-cell subsets were relatively depleted (Figs. 7B-C). Representative CT images together with matched BALF single-cell profiles further illustrated the distinct pulmonary cellular landscapes across model-defined risk groups (Fig. 7D).

GSEA demonstrated coordinated transcriptional reprogramming across major pulmonary cell populations in high-risk ARDS. Neutrophils, T cells, and NK cells showed enrichment of multiple intracellular signalling pathways, including mTOR-, Wnt-, phosphatidylinositol-, AMPK-, and MAPK-related programs, alongside selected metabolic pathways, whereas ribosomal, spliceosomal, proteasomal, and oxidative phosphorylation programs were broadly suppressed (Fig. 7E). Macrophages additionally exhibited attenuation of cytokine, glycolytic, adhesion, and innate immune pathways, while epithelial cells showed prominent activation of NF-κB-, MAPK-, TNF-, and stress-response signalling with concomitant suppression of metabolic and cellular homeostatic programs (**Figure S9**).

## Discussion

In this multicentre study, we developed and externally validated a multimodal framework integrating clinical, quantitative CT, and serum proteomic data for ARDS risk stratification. Model-defined risk states were associated with distinct biological features, including innate immune dysregulation, epithelial and vascular injury, impaired cellular homeostasis, and neutrophil-dominant pulmonary reprogramming. Exploratory time-varying Cox further suggested heterogeneity in glucocorticoid-associated outcomes across risk groups, supporting the potential relevance of multimodal risk profiling for treatment stratification.

To date, several studies have applied omics-based approaches to outcome prediction in ARDS and related critical illness. Spottiswoode *et al.* reported an AUC of 0.79 for an integrated host-transcriptomic and plasma-metagenomic classifier of in-hospital mortality in sepsis, compared with 0.75 for host transcriptomics alone and 0.69 for APACHE III [18]. Although the addition of microbial information did not significantly outperform the host-only model, proteomic integration in our study consistently improved discrimination beyond both the clinical-imaging and proteomic models. This supports the complementary value of combining structural lung injury, physiological severity, and circulating host-response signals for risk stratification in ARDS. Importantly, the clinical-imaging model maintained prognostic performance across independent cohorts, while incorporation of serum proteomic information provided additional improvement. This is particularly relevant in critical illness, where model performance may vary across centers and ICU populations[13]. By combining clinical, structural, and molecular information, the multimodal framework may better capture the biological heterogeneity underlying ARDS outcomes. Its relevance also extended beyond mortality: higher model-derived risk was associated with adverse

early respiratory-support trajectories, including lower likelihood of early extubation success and greater risk of non-invasive mechanical ventilation failure. These findings complement previous studies linking molecular signatures, including the respiratory microbiome, to prolonged mechanical ventilation[19] and suggest that multimodal risk stratification may provide information across multiple clinically relevant outcomes.

Proteomic profiling provided biological context beyond the clinical-imaging-defined risk phenotype. The high-risk state was characterized by convergent signals of epithelial injury and aberrant repair, immune dysregulation, and extrapulmonary stress. Elevated KRT19[20] and ITGB6[21] supported involvement of the injured epithelial compartment, while GDF15 reflected broader cellular stress[22]. Alterations in IL6, IL10, and IL1RL1 further suggested concurrent inflammatory activation and counter-regulation. The concordant association of proteomic NPPB and clinically measured NT-proBNP with increasing risk also indicated a contribution from cardiovascular stress[23]. Together, these findings suggest that proteomics captures biological processes not fully represented by physiological severity or structural imaging, helping explain its incremental value in multimodal risk prediction.

Corticosteroid treatment was associated with lower mortality in the high-risk group, suggesting clinically relevant treatment heterogeneity. This group showed activation of inflammatory and innate immune pathways, including IL-6, NF-κB, TREM1, Toll-like receptor, HMGB1, and neutrophil-degranulation signalling, together with reduced PPAR and LXR/RXR activity. This pattern is consistent with persistent innate immune activation and impaired regulatory or resolution programs, providing a biological context for the observed treatment-associated heterogeneity. Similar corticosteroid heterogeneity has been reported across biologically defined ARDS phenotypes[24]. However, given the observational treatment allocation and potential time-dependent confounding, these findings remain hypothesis-generating and require prospective validation before guiding treatment selection.

The overlap in proteomic signatures between clinical-imaging and multimodal risk groups suggests that clinical and CT features capture part of the underlying biological gradient, while proteomics provides additional molecular resolution. This supports a potential tiered strategy in which clinical-imaging assessment serves as an accessible first step, with proteomic profiling added when greater biological characterization is needed. Whether this incremental gain justifies the additional cost, infrastructure, and turnaround time will require prospective decision-impact and cost-effectiveness evaluation.

BALF single-cell profiling provided complementary biological context for the multimodal risk groups. High-risk patients showed marked neutrophil expansion with relative depletion of macrophage and lymphoid populations, consistent with a myeloid-dominant immune state. Similar neutrophil programs have been reported in severe pneumonia and sepsis, where inflammatory and immunoregulatory neutrophil states coexist with impaired adaptive immunity[25, 26]. Systemic single-cell studies further link emergency granulopoiesis and immature neutrophil states to poor-outcome sepsis phenotypes and impaired T-cell activation[27]. Together, these findings suggest that high-risk ARDS may involve maladaptive myeloid reprogramming rather than simply greater inflammatory burden. The accompanying suppression of ribosomal, RNA-processing, proteostatic, and metabolic programs further indicates impaired cellular homeostasis. The convergence of these pulmonary cellular alterations with serum proteomic abnormalities and CT-defined lung injury suggests that the multimodal score captures an integrated state of structural injury, host-response dysregulation, and impaired homeostatic capacity beyond conventional measures of disease severity.

The reader study provided a complementary assessment of AI-assisted prognostication in ARDS. The clinical-imaging model showed higher discrimination than physician assessment and a general-purpose large language

model, suggesting that structured integration of clinical and imaging information may provide more consistent quantitative risk estimates. However, such models are intended to complement rather than replace clinical judgement, which incorporates longitudinal trajectories, treatment response, and contextual factors beyond baseline prediction. Prospective studies are needed to determine whether model-assisted assessment improves clinician risk estimation, decision-making, and ultimately patient outcomes.

This study has several strengths. It included a large multicentre population with centre-separated model development, internal testing, and external validation across 10 hospitals. The tiered design enabled evaluation of clinically accessible models with and without proteomic information, while performance was examined across mortality horizons, clinically relevant subgroups, respiratory-support outcomes, and a reader study. Biological characterization further extended from circulating proteins and pathways to pulmonary cellular states, providing complementary evidence that model-defined risk reflects clinically and biologically meaningful heterogeneity rather than cohort-specific statistical associations.

Several limitations should be acknowledged. First, the glucocorticoid analyses were observational. Although time-varying Cox models accounted for treatment timing and measured baseline covariates, residual confounding by indication, particularly time-varying confounding related to evolving disease severity, and heterogeneity in glucocorticoid regimens cannot be excluded; these findings therefore require prospective confirmation. Second, BALF single-cell analysis was exploratory and included only 26 patients, limiting inference for less abundant cell populations and subgroup-specific transcriptional programs. Third, inclusion required an evaluable index CT within the prespecified window, which may have introduced CT-based selection and survivor bias and may limit generalizability to patients without early CT imaging. Finally, the models were based primarily on measurements around the index CT and did not capture longitudinal clinical, imaging, or molecular changes; the reliance on manual lung segmentation may also limit immediate scalability in routine practice.

In conclusion, integration of clinical, imaging, and proteome enabled robust prognostic stratification of ARDS and identified risk states with distinct respiratory trajectories and biological profiles. The observed heterogeneity in glucocorticoid-associated outcomes further suggests that multimodal risk profiling may provide a framework linking prognostic stratification with biologically informed treatment stratification, warranting prospective evaluation.

## Methods

# Patient population and study design

The overall study design is shown in Fig. 1. This multicentre study combined retrospective clinical-imaging cohorts with prospectively collected serum proteomic and BALF single-cell transcriptomic cohorts. Consecutive patients meeting the Global Definition of ARDS[28] were screened at 10 medical centres in China between June 2021 and October 2025. Dataset 1 included 2,041 patients and was used for model development and internal testing. Patients from Zhongshan Hospital and Shanghai General Hospital constituted the training cohort (n = 1,715), whereas patients from Guangdong Provincial People's Hospital and Renmin Hospital of Wuhan University constituted the internal test cohort (n = 326). Dataset 2 included 827 patients from six additional centres and served as the external validation cohort (**Supplementary Methods**).

Serum samples were prospectively collected from 879 patients with corresponding clinical and imaging data, including 554 in the training cohort, 176 in the internal test cohort, and 149 in the external validation cohort

(Fig. 1A). Among these patients, a nested subset of 26 patients also underwent BALF sampling and scRNA-seq for exploratory characterization of pulmonary cellular states associated with multimodal model-defined risk groups.

# Baseline clinical parameters and clinical data pre-processing

The index CT (T0) was defined as the first evaluable chest CT obtained within 48 h after ARDS diagnosis. Patients without an evaluable chest CT within this prespecified window were excluded. Baseline variables were defined using information available by index CT acquisition. Time-varying physiological, laboratory, arterial blood-gas, and respiratory variables were represented by the most recent value within the preceding 24 h, while cumulative measures such as urine output were assessed over the same period. Demographics and comorbidities were determined by T0.

A total of 75 candidate clinical variables were collected, encompassing demographics, comorbidities, ARDS etiology, vital signs, respiratory-support parameters, arterial blood-gas measurements, routine hematological and biochemical indices, and measures of organ dysfunction (**Table S5**). Information on ventilatory support and systemic glucocorticoid exposure was additionally recorded for secondary outcome and treatment analyses. Variables with > 15% missingness in the training cohort were excluded. Variables with ≤ 15% missingness were imputed using multivariate imputation by chained equations (MICE)[29] under a missing-at-random assumption conditional on observed covariates[29]. Imputation was performed independently within the training, internal-test, and external-validation cohorts.

# Clinical outcomes

The primary outcome was all-cause mortality within 30 days after CT acquisition. The prognostic performance of the original 30-day mortality model at longer follow-up horizons was additionally assessed for 60-, 90-, and 180-day mortality using time-dependent ROC analysis; the model was not refitted for these secondary analyses.

# Multimodal model development

The model-development workflow is shown in Fig. 1B. Whole-lung regions, including aerated and consolidated parenchyma, were first automatically segmented by our previous AI model on chest CT scans[30], followed by manual review and correction by two thoracic radiologists. Disagreements were resolved by consensus or adjudication by a senior radiologist. Inter-annotator agreement after AI-assisted refinement was high (median Dice similarity coefficient, 0.92 [IQR, 0.88–0.95]). The reference standard was 30-day all-cause mortality from the index CT, ascertained from electronic health records, ICU databases, and follow-up interviews and independently adjudicated by investigators blinded to model outputs.

For the clinical model, 75 candidate variables underwent univariate screening followed by LASSO-based selection, and 30 retained predictors were subsequently entered into multivariable logistic regression to generate patient-level mortality probabilities. For the radiomics model, 104 quantitative features were extracted from annotated lung regions; after reproducibility assessment, redundancy reduction, and LASSO selection, 25 features were retained and entered into multivariable logistic regression to derive radiomics mortality probability. ARDS-Net, a ResNet-50-based deep-learning model, was independently trained on lung-masked CT volumes resampled to 1 × 1 × 3 mm

voxel spacing and cropped to 256 × 256 × 128 voxels. For patients with paired serum proteomic data, proteins underwent quality control, batch correction, minimum redundancy maximum relevance (mRMR) pre-filtering, and LASSO selection, yielding 22 features that were subsequently entered into a multivariable logistic-regression model to generate proteomic mortality probabilities (**Supplementary Methods**).

Two separate model-level fusion models were developed. The clinical-imaging model integrated predicted probabilities from the clinical, radiomics, and deep-learning models, whereas the multimodal model additionally incorporated the proteomic model. Component-model probabilities were standardized using training-set parameters. Five candidate meta-classifiers were compared during model development, and L2-penalized logistic regression was selected as the final fusion method based on discrimination, stability, and parsimony (**Figure S10** and **Supplementary Methods**). All model-development procedures were confined to the training data, and the finalized models were applied unchanged to the internal-test and external-validation cohorts. Detailed methods are provided in the **Supplementary Methods** and **Tables S6-S9**.

$$\eta_{\,\mathrm{clinical-imaging}} = -0.9231 + 0.1168 Z_{\mathrm{radiomics}} + 1.0532 Z_{\mathrm{clinical}} + 1.0322 Z_{\mathrm{deep\ learning}}$$

whereas the multimodal predictor was

$$\eta_{\,\mathrm{multimodal}} = -0.7201 + 0.1106 Z_{\mathrm{radiomics}} + 0.4794 Z_{\mathrm{clinical}} + 0.6635 Z_{\mathrm{deep\ learning}} + 0.7981 Z_{\mathrm{proteomic}}$$

Final 30-day mortality probabilities were calculated as $P = 1/[1 + \exp(-\eta\ )]$. After model finalization, all preprocessing parameters and fusion-model coefficients were locked and applied unchanged to the internal-test and external-validation cohorts.

# Model assessment

Model performance was primarily assessed using the receiver operating characteristic (ROC) curves and quantified by the area under the ROC curve (AUROC) with 95% confidence intervals (CI) and confusion matrix, with additional evaluation metrics including sensitivity, specificity, positive predictive value, negative predictive value, and accuracy across all cohorts. Calibration was evaluated using calibration plots, Brier scores, calibration-in-the-large, and calibration slopes. Decision curve analysis was used to quantify net benefit across a range of clinically relevant threshold probabilities. Differences in discrimination were quantified as ΔAUROC, with 95% CIs and two-sided *P* values estimated using 2,000 paired patient-level bootstrap replicates. Furthermore, model performance was evaluated in clinically relevant subgroups defined by age (< 65 versus ≥ 65 years), ARDS etiology (direct versus indirect), and the presence or absence of chronic pulmonary disease. AUCs with 95% CIs were calculated separately for the clinical-imaging, proteomic and fully integrated multimodal models.

# Reader study

We conducted a multi-reader, multi-case observer study using all 326 patients from the internal test cohort. Clinicians first reviewed complete chest CT DICOM volumes without access to radiology reports, treatment information, or outcomes. In a second session, they reassessed the same cases after receiving clinical information available at the index CT time point (**Table S5**). For each session, readers provided a continuous 30-day mortality probability from 0% to 100% and a 5-point Likert risk score; the continuous probability was used for the primary comparison with the clinical-imaging model. ChatGPT-5.4 (OpenAI, San Francisco, CA, USA) independently

evaluated the same 326 cases using a standardized prompt (**Supplementary Methods**) and the same index-time clinical information. Each case was assessed independently, without access to patient outcomes, clinical-imaging model predictions, or clinician ratings. ChatGPT was instructed to generate a 0–100% 30-day mortality probability, which was recorded without post hoc modification.

## Association with subsequent respiratory-support outcomes

To assess whether the mortality risk score captured subsequent respiratory trajectories, index CT acquisition (T0) was used as the common time origin. NIMV failure was defined as escalation to IMV or death within 7 days after T0 among patients receiving NIMV at T0. Early successful weaning was defined as liberation from IMV within 7 days after T0 without death or reintubation during the subsequent 7 days among patients receiving IMV at T0. Associations of the clinical-imaging and multimodal risk scores with these outcomes were evaluated using ROC analysis and restricted cubic-spline logistic regression.

## Risk stratification and treatment-associated outcomes

Patients were stratified into low-, intermediate-, and high-risk groups using the 33rd and 67th percentiles of predicted probability derived from the training cohort, with these thresholds applied unchanged to the internal-test and external-validation cohorts. Systemic glucocorticoid initiation within 7 days after T0 was modelled as a time-varying exposure; patients contributed unexposed person-time until initiation and exposed person-time thereafter, whereas those without initiation within 7 days remained unexposed. Associations with 30-day mortality and treatment-by-risk interactions were evaluated using multivariable time-dependent Cox models (**Supplementary Methods**).

## Proteomic data analysis

Serum samples paired with the clinical assessment were prospectively collected at the time of chest CT acquisition. Protein abundance was quantified using the Olink Inflammation and Cardiovascular Explore panels. A total of 711 proteins were included in the analysis, covering a broad range of biological processes, including inflammation, immune regulation, cardiovascular function, and metabolism (**Supplementary Methods)**. Differentially abundant proteins were identified using a Benjamini-Hochberg adjusted *P* value < 0.05 and an absolute fold change ≥ 1.3 using the limma package in *R* [31]. Significant proteins were subsequently subjected to Ingenuity Pathway Analysis (IPA; QIAGEN Inc., Redwood City, CA, USA) to identify enriched canonical pathways.

## Single-cell transcriptomic data analysis

BALF samples were collected on the same day as the index chest CT (T0) in all patients included in the single-cell cohort. Single-cell RNA sequencing data from BALF specimens of ARDS were analyzed to delineate the cellular ecosystem associated with multimodal risk groups. After quality control, cells from multiple samples were integrated, clustered and annotated into major immune and epithelial populations using a standard Seurat-based pipeline. Details in **Supplementary Methods**.

# Statistical analysis

Survival probabilities were assessed with Kaplan-Meier survival analysis and log-rank tests. A two-tailed p-value < 0.05 was considered statistically significant. All statistical analyses were performed using *R* (version 4.2.1) and *Python* (version 3.6). Radiomic feature extraction, feature selection, and model development were performed using the uAI Research Portal (Shanghai United Imaging Intelligence Co., Ltd., China; https://urp.united-imaging.com) [32]. All input features from each modality were normalized before model development.

# Declarations

### Author contributions

Z. Song was responsible for study conception and design. M. Lin, S. Lu, and Y. Zeng performed all data analyses and wrote the manuscript. N. Zheng, Y. Chen, C. Chen, X. Zou, X. Tian, F. Xu, Z. Chen, S. Wu, C. Gao, X. Wu, Z. Zhang, Q. Liu, Y. Shen, Y. Lin, Y. Deng, R. Ji, J. Song, L. Wang, C. Peng, and X. Tian contributed to patient recruitment, clinical data collection, biological sample acquisition, and cohort management across participating centres. M. Lin, S. Lu, and Y. Zeng contributed to imaging, proteomic, and single-cell transcriptomic analyses. R. Wang, L. Chen, L. Zhan, and Z. Song provided clinical and methodological supervision, contributed to interpretation of the findings, and critically revised the manuscript for important intellectual content. Z. Song supervised the overall study and obtained funding. All authors reviewed and approved the final manuscript and accept responsibility for the integrity of the work.

### Funding

This work was supported by the Prevention and Control of Emerging and Major Infectious Diseases-National Science and Technology Major Project (2025ZD01902600), the Science and Technology of Shanghai Committee (25TS1411600 and 23Y31900100), Shanghai Municipal Health Commission (2023ZDFC0101), and the National Natural Science Foundation of China (82604339).

### Conflicts of interest

Authors declare no conflicts of interest related to the present study.

### Data availability

The proteome data was deposited at OMIX, China National Center for Bioinformation (https://ngdc.cncb.ac.cn/omix/release/OMIX020688). Processed single-cell RNA sequencing data, including the annotated Seurat object and associated metadata, have been deposited in Figshare. Clinical data support the findings of this study are available from the corresponding author (Z. Song, song.zhenju@zs-hospital.sh.cn), upon reasonable request.

### Code availability

The core code for the research can be found on the corresponding author's GitHub page (https://github.com/MengnaLin5251/multimodal_ARDS).

The study was approved by the Ethics Committee of Zhongshan Hospital, Fudan University (Ethics Approval No. B2022-175).

## References


1. Bos LDJ, Ware LB (2022) Acute respiratory distress syndrome: causes, pathophysiology, and phenotypes. Lancet 400(10358):1145–1156
2. Beitler JR, Thompson BT, Baron RM, Bastarache JA, Denlinger LC, Esserman L et al (2022) Advancing precision medicine for acute respiratory distress syndrome. Lancet Respir Med 10(1):107–120
3. Nasa P, Bos LD, Estenssoro E, van Haren FMP, Neto AS, Rocco PRM et al (2025) Defining and subphenotyping ARDS: insights from an international Delphi expert panel. Lancet Respir Med 13(7):638–650
4. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A et al (2016) Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. JAMA 315(8):788–800
5. Moor M, Bennett N, Plecko D, Horn M, Rieck B, Meinshausen N et al (2023) Predicting sepsis using deep learning across international sites: a retrospective development and validation study. EClinicalMedicine 62:102124
6. Luong CQ, Dao CX, Nguyen MH, Pham DT, Pham QT, Vu TT et al (2025) Prognostic value of the PaO(2)/FiO(2) ratio for mortality in acute respiratory distress syndrome: a retrospective observational study in a lower-middle-income country. BMJ Open 15(12):e100589
7. Bennett RM, Housel KC, Jones TK, Miano TA, Emre G, Turner A et al (2026) Epidemiology of the acute respiratory distress syndrome and the prognostic validity of SpO(2):FiO(2) under the expanded Global definition. Crit Care 30(1)
8. Bedoya AD, Futoma J, Clement ME, Corey K, Brajer N, Lin A et al (2020) Machine learning for early detection of sepsis: an internal and temporal validation study. JAMIA Open 3(2):252–260
9. Leonardi A, Scipione R, Alfieri G, Petrillo R, Dolciami M, Ciccarelli F et al (2020) Role of computed tomography in predicting critical disease in patients with covid-19 pneumonia: A retrospective study using a semiautomatic quantitative method. Eur J Radiol 130:109202
10. Zhou Y, Mei S, Wang J, Xu Q, Zhang Z, Qin S et al (2024) Development and validation of a deep learning-based framework for automated lung CT segmentation and acute respiratory distress syndrome prediction: a multicenter cohort study. EClinicalMedicine 75:102772
11. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X et al (2020) CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology 295(1):202–207
12. Yu AC, Mohajer B, Eng J (2022) External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. Radiol Artif Intell 4(3):e210064
13. Wang S, Liu X, Yuan S, Bian Y, Wu H, Ye Q (2025) Artificial intelligence based multispecialty mortality prediction models for septic shock in a multicenter retrospective study. NPJ Digit Med 8(1):228
14. Bitker L, Talmor D, Richard JC (2022) Imaging the acute respiratory distress syndrome: past, present and future. Intensive Care Med 48(8):995–1008
15. Sarma A, Christenson SA, Shoshana BZ, Oliveira AP, Neyton LPA, Mick E et al (2025) Acute Respiratory Distress Syndrome Molecular Phenotypes Have Distinct Lower Respiratory Tract Transcriptomes. Am J Respir Crit Care Med 211(12):2352–2362
16. Lin M, Xu F, Sun J, Song J, Shen Y, Lu S et al (2024) Integrative multi-omics analysis unravels the host response landscape and reveals a serum protein panel for early prognosis prediction for ARDS. Crit Care 28(1):213

17. Kim C, Gadgil SU, Lee SI (2026) Transparency of medical artificial intelligence systems. Nat Rev Bioeng 4(1):11–29
18. Spottiswoode N, Neyton LP, Mick E, Kalantar KL, Hao S, Lydon EC et al (2026) Host-Microbe Multiomic Profiling Predicts Mortality in Sepsis. Am J Respir Crit Care Med 212(1):95–104
19. Montassier E, Kitsios GD, Radder JE, Le Bastard Q, Kelly BJ, Panzer A et al (2023) Robust airway microbiome signatures in acute respiratory failure and hospital-acquired pneumonia. Nat Med 29(11):2793–2804
20. Stern JB, Paugam C, Validire P, Adle-Biassette H, Jaffre S, Dehoux M et al (2006) Cytokeratin 19 fragments in patients with acute lung injury: a preliminary observation. Intensive Care Med 32(6):910–914
21. Sheppard D (2015) Epithelial-mesenchymal interactions in fibrosis and repair. Transforming growth factor-beta activation by epithelial cells and fibroblasts. Ann Am Thorac Soc 12(Suppl 1):S21–23
22. Clark BJ, Bull TM, Benson AB, Stream AR, Macht M, Gaydos J et al (2013) Growth differentiation factor-15 and prognosis in acute respiratory distress syndrome: a retrospective cohort study. Crit Care 17(3):R92
23. Bajwa EK, Januzzi JL, Gong MN, Thompson BT, Christiani DC (2008) Prognostic value of plasma N-terminal probrain natriuretic peptide levels in the acute respiratory distress syndrome. Crit Care Med 36(8):2322–2327
24. Lin M, Xu F, Deng Y, Wei Y, Shi F, Xie Y et al (2026) Large-scale proteomic profiling identifies distinct inflammatory phenotypes in acute respiratory distress syndrome: a multicentre, prospective cohort study. Eur Respir J 67(2)
25. Xiao K, Cao Y, Han Z, Zhang Y, Luu LDW, Chen L et al (2025) A pan-immune panorama of bacterial pneumonia revealed by a large-scale single-cell transcriptome atlas. Signal Transduct Target Ther 10(1):5
26. Shen R, Jiang Y, Liu G, Gao S, Sun H, Wu X et al (2025) Single-Cell Landscape of Bronchoalveolar Lavage Fluid Identifies Specific Neutrophils during Septic Immunosuppression. Adv Sci (Weinh) 12(11):e2406218
27. Kwok AJ, Allcock A, Ferreira RC, Cano-Gamez E, Smee M, Burnham KL et al (2023) Neutrophils and emergency granulopoiesis drive immune suppression and an extreme response endotype during sepsis. Nat Immunol 24(5):767–779
28. Matthay MA, Arabi Y, Arroliga AC, Bernard G, Bersten AD, Brochard LJ et al (2024) A New Global Definition of Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med 209(1):37–47
29. Azur MJ, Stuart EA, Frangakis C, Leaf PJ (2011) Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 20(1):40–49
30. Wang C, Shao J, He Y, Wu J, Liu X, Yang L et al (2024) Data-driven risk stratification and precision management of pulmonary nodules detected on chest computed tomography. Nat Med 30(11):3184–3195
31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43(7):e47
32. Wu J, Xia Y, Wang X, Wei Y, Liu A, Innanje A et al (2023) uRP: An integrated research platform for one-stop analysis of medical images. Front Radiol 3:1153784

## Figures

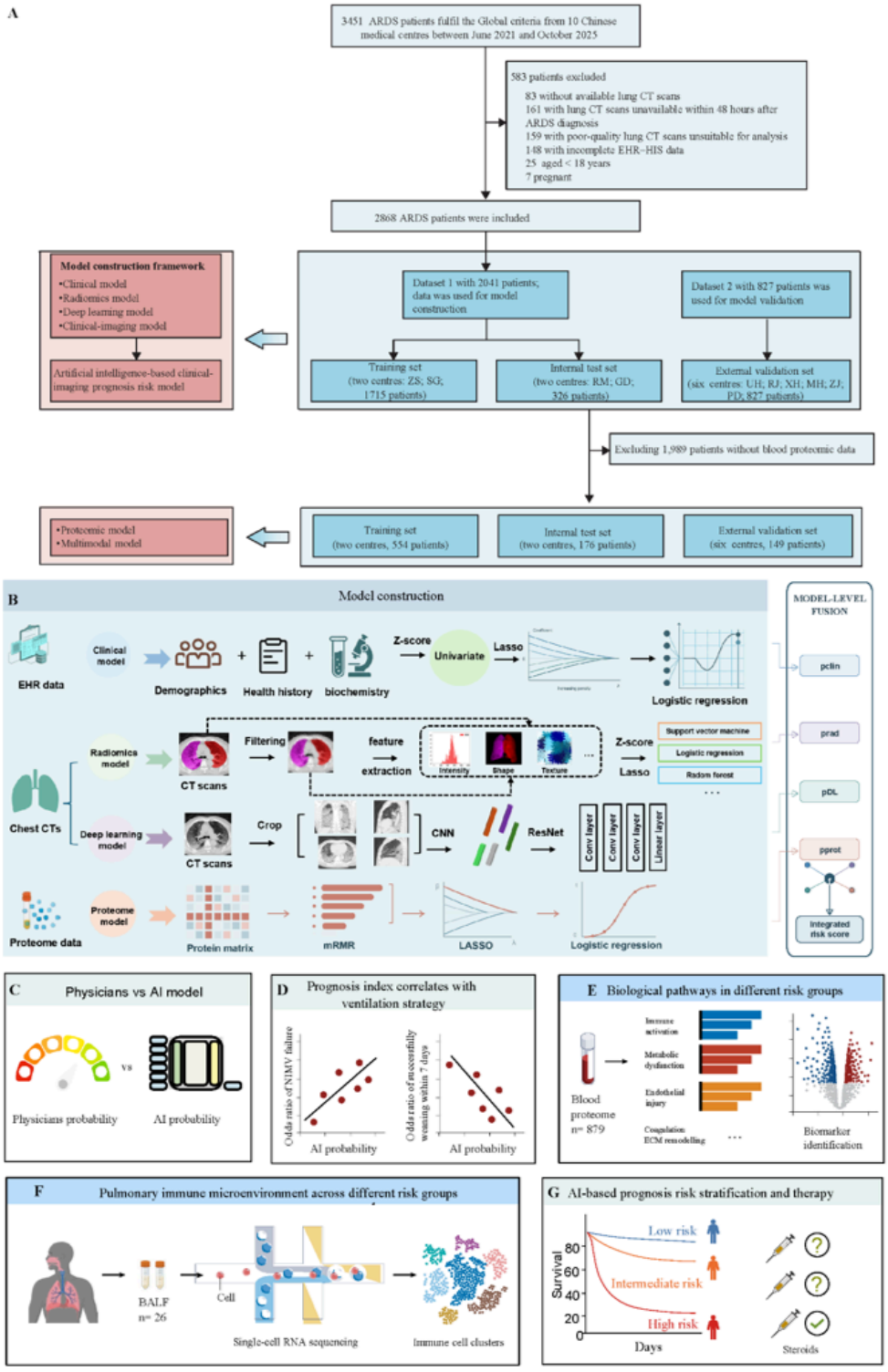
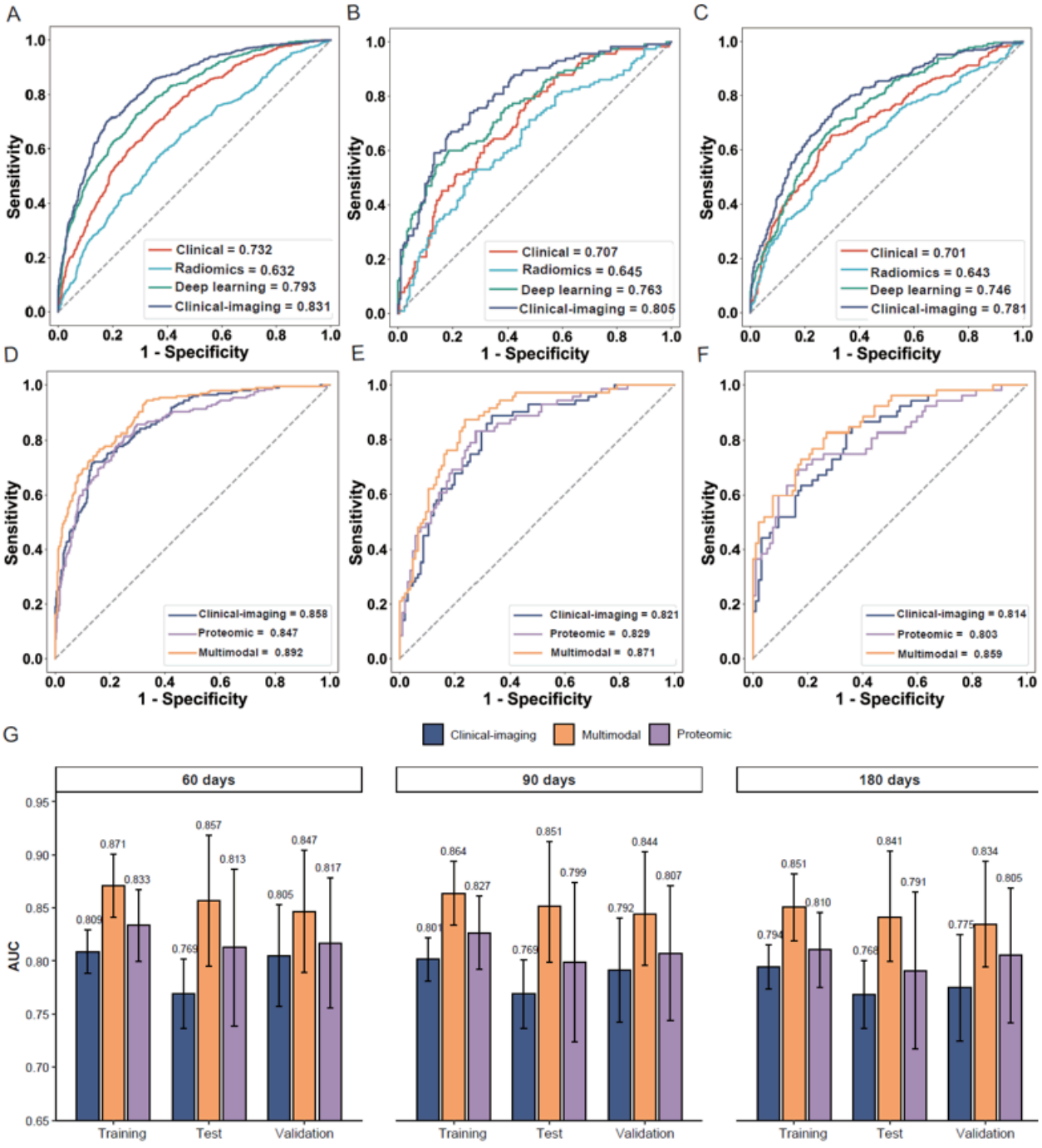


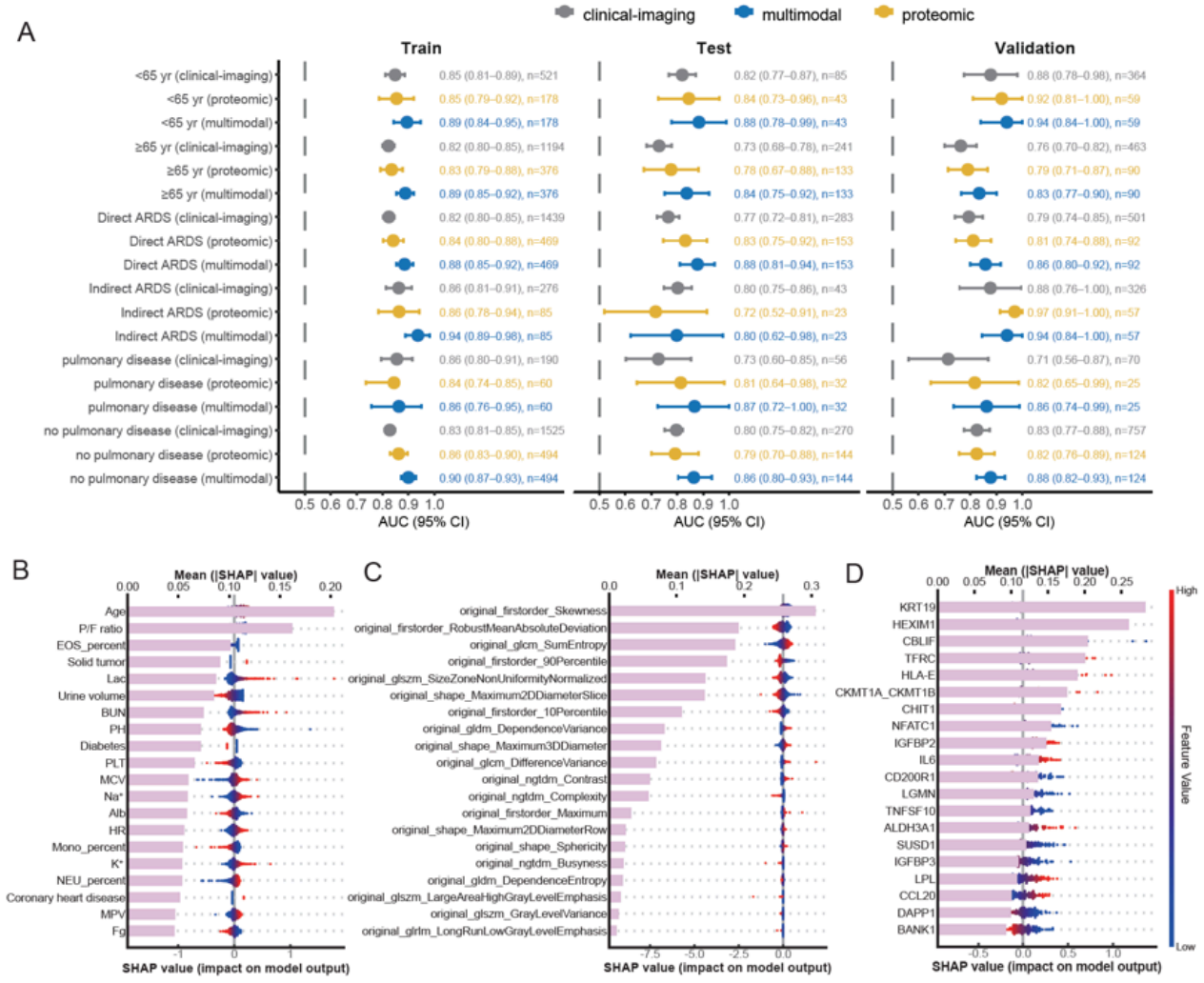

**Figure 1**

**Study flowchart.** ZS: Zhongshan Hospital, Fudan University; SG: Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine; RM: Renmin Hospital of Wuhan University; GD: Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences; UH: Union Hospital, Tongji Medical College, Huazhong University of Science and Technology; RJ: Renji Hospital, Shanghai Jiao Tong University School of Medicine; XH: Xinhua Hospital, Shanghai Jiao Tong University School of Medicine; MH: Minhang Hospital, Fudan University; ZJ: the Second Affiliated Hospital, Zhejiang University School of Medicine; PD: Pudong Hospital, Fudan University.




## Figure 2

**Predictive performance of different models for mortality across the training, testing, and validation cohorts.** (**A-C**) ROC curves of the clinical, radiomics, deep learning, and clinical-imaging models in the training (A), internal test (B), and validation (C) cohorts, respectively. The AUC for each model is shown in the corresponding panel. (**D-F**) ROC curves of the clinical-imaging, proteomic, and multimodal models in the training (D), testing (E), and validation (F) cohorts, respectively. (**G**) Time-dependent AUCs of the clinical-imaging, proteomic, and multimodal models for predicting 60-day, 90-day, and 180-day mortality in the training, testing, and validation cohorts. Bars represent AUC values, and error bars indicate 95% CIs.




## Figure 3

**Model interpretability, comparative prognostic performance, and subgroup analyses.** (**A**) Robustness of model discrimination across clinically relevant subgroups. AUCs with 95% confidence intervals are shown for the proteomic, clinical-imaging and multimodal models, stratified by age (<65 vs. ≥65 years), ARDS etiology (direct vs. indirect), and the presence or absence of pre-existing pulmonary disease. **(B–D)**SHAP (SHapley Additive exPlanations) summary plots showing feature importance and direction of contribution for the clinical model (A), radiomics model (B), and proteomic model (C). Features are ranked by mean absolute SHAP value. Each point represents an individual participant; color indicates the corresponding feature value, with red denoting higher values and blue denoting lower values. Positive and negative SHAP values indicate contributions toward higher and lower model-predicted risk, respectively. **Abbreviations:** P/F ratio, ratio of arterial partial pressure of oxygen to fraction of inspired oxygen ($PaO_2/FiO_2$); EOS_percent, eosinophil percentage; Lac, lactate; BUN, blood urea nitrogen; pH, potential of hydrogen; PLT, platelet count; MCV, mean corpuscular volume; $Na^+$, sodium; Alb, albumin; HR, heart rate; Mono_percent, monocyte percentage; $K^+$, potassium; NEU_percent, neutrophil percentage; MPV, mean platelet volume; Fg, fibrinogen.

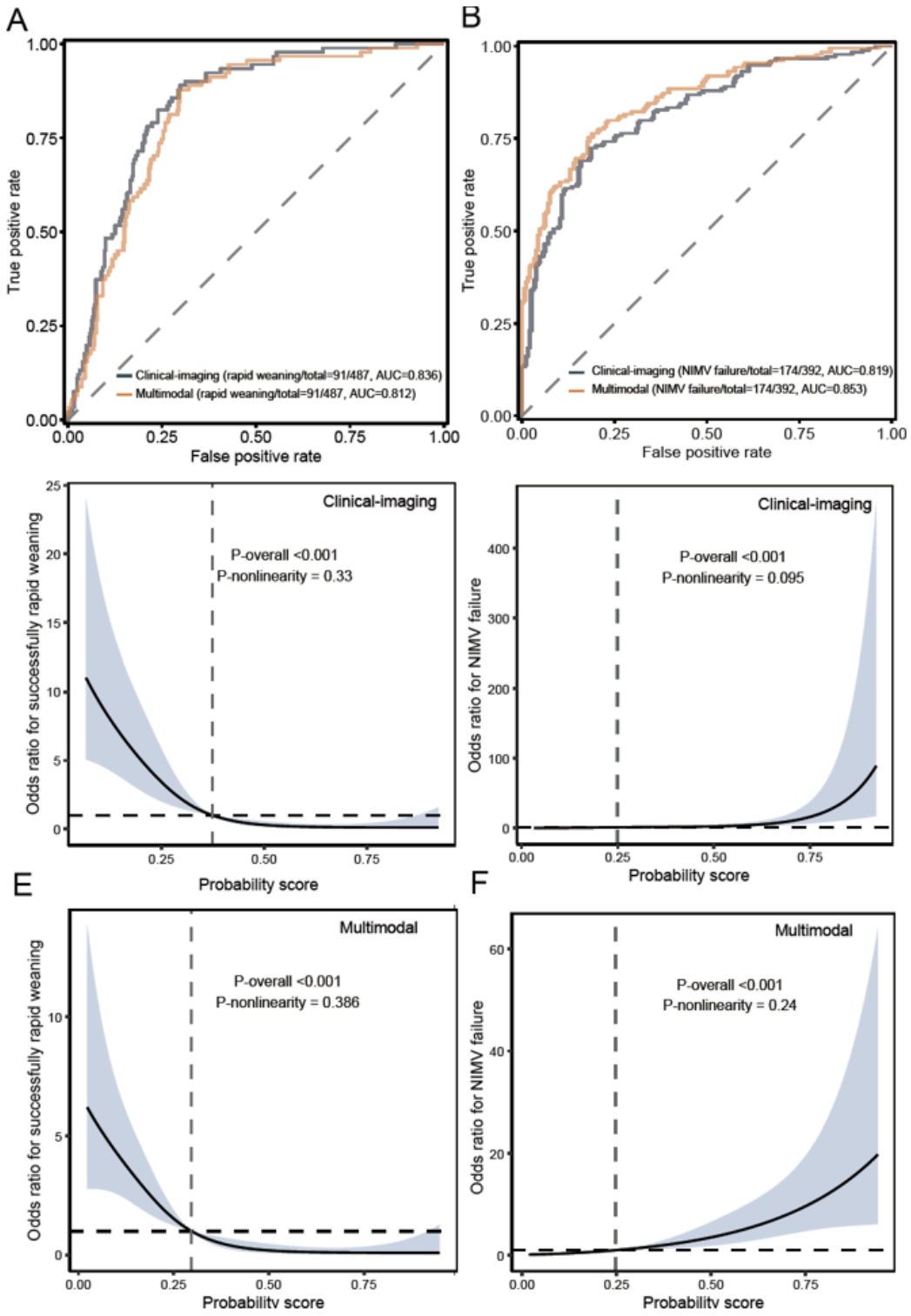
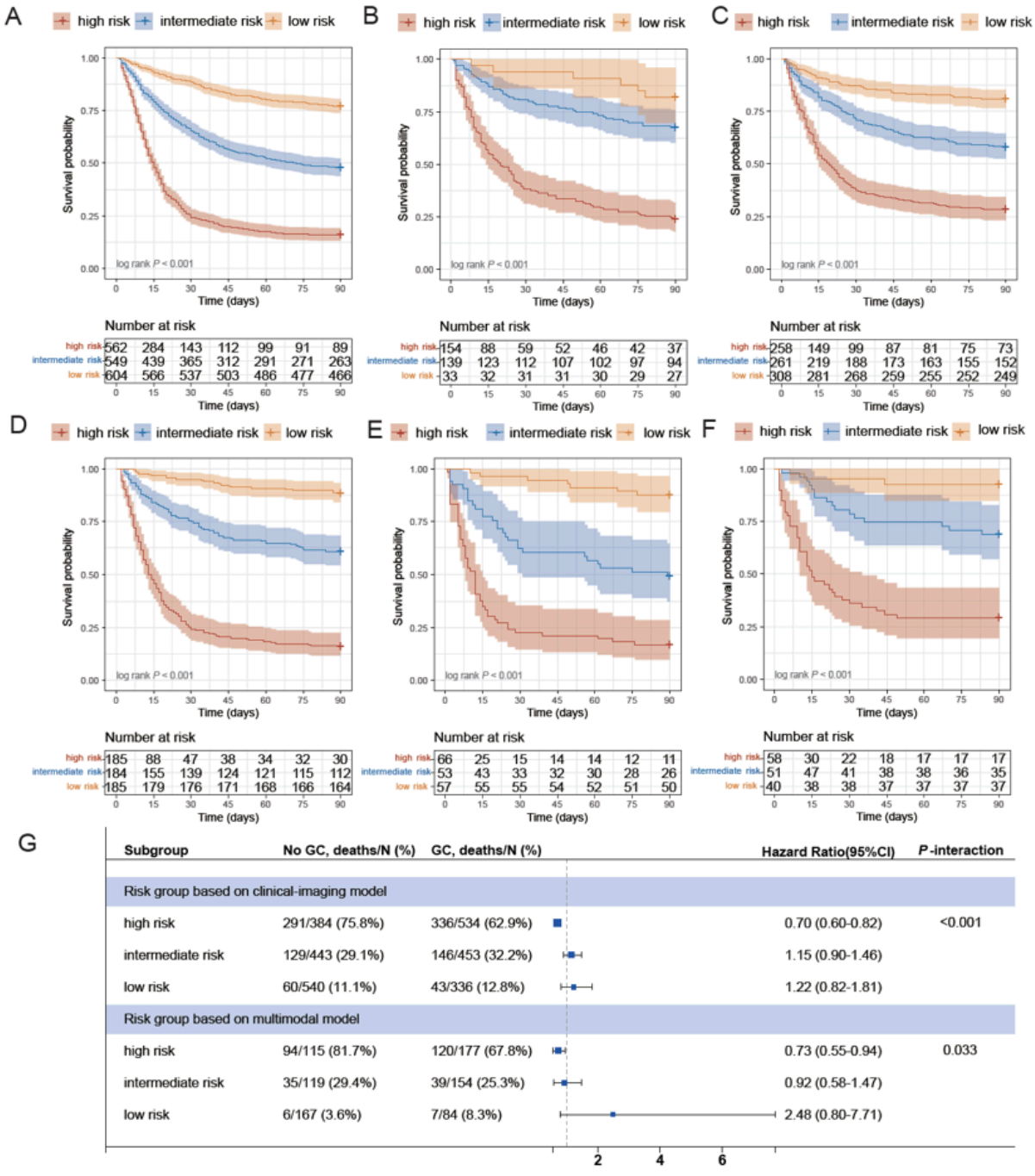



**Figure 4**

**Association of model-derived prognostic risk with respiratory-support outcomes in the matched proteomic cohort.** (**A-B**) ROC curves comparing the clinical-imaging and multimodal models for predicting successful rapid weaning from invasive mechanical ventilation within 7 days (A) and non-invasive mechanical ventilation (NIMV) failure (B) among patients with matched proteomic data. Event counts, total sample sizes, and areas under the ROC curve (AUCs) are shown in the corresponding panels. The diagonal dashed line indicates chance-level discrimination. (**C-D**) Restricted cubic-spline analyses showing the associations between the clinical-imaging

model-predicted mortality probability and the odds of successful rapid weaning (C) and NIMV failure (D). (**E-F**) Corresponding restricted cubic-spline analyses for the multimodal model, which additionally incorporated proteomic information, for successful rapid weaning (E) and NIMV failure (F). Solid black lines represent estimated odds ratios (ORs), and shaded areas indicate 95% confidence intervals. Horizontal dashed lines indicate an OR of 1, and vertical dashed lines indicate the model-specific predicted probability at which the estimated OR equals 1. Overall and non-linearity *P* values are shown for each spline analysis.




## Figure 5

**Risk stratification and heterogeneity in outcomes associated with glucocorticoid treatment.** (**A-C**) Kaplan–Meier survival curves for the high-, intermediate-, and low-risk groups defined by tertiles of the predicted probability from

the clinical-imaging model in the training (A), internal test (B), and validation (C) cohorts. Shaded areas represent 95% CIs, and numbers at risk are shown below each panel. Differences among risk groups were assessed using the log-rank test. (**D-F**) Corresponding Kaplan-Meier survival curves for the high-, intermediate-, and low-risk groups defined by tertiles of the predicted probability from the multimodal model in the training (D), internal test (E), and validation (F) cohorts. (**G)** Forest plot showing the association between time-varying systemic glucocorticoid initiation and 30-day mortality across high-, intermediate-, and low-risk groups defined by the clinical-imaging and multimodal models. Hazard ratios (HRs) and 95% confidence intervals were estimated using multivariable time-dependent Cox proportional-hazards models. Global *P* values for interaction assess heterogeneity in glucocorticoid-associated outcomes across risk groups.

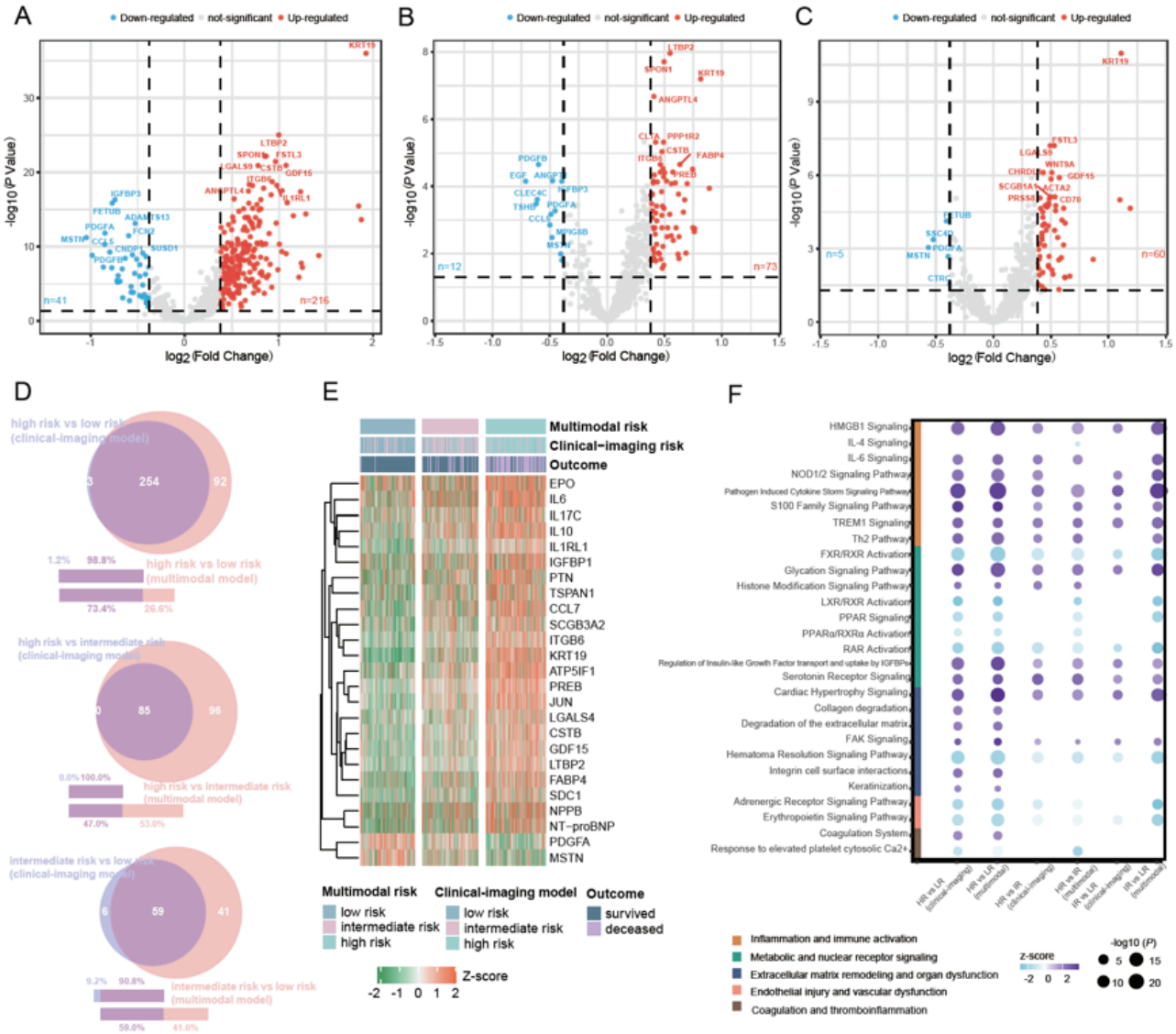
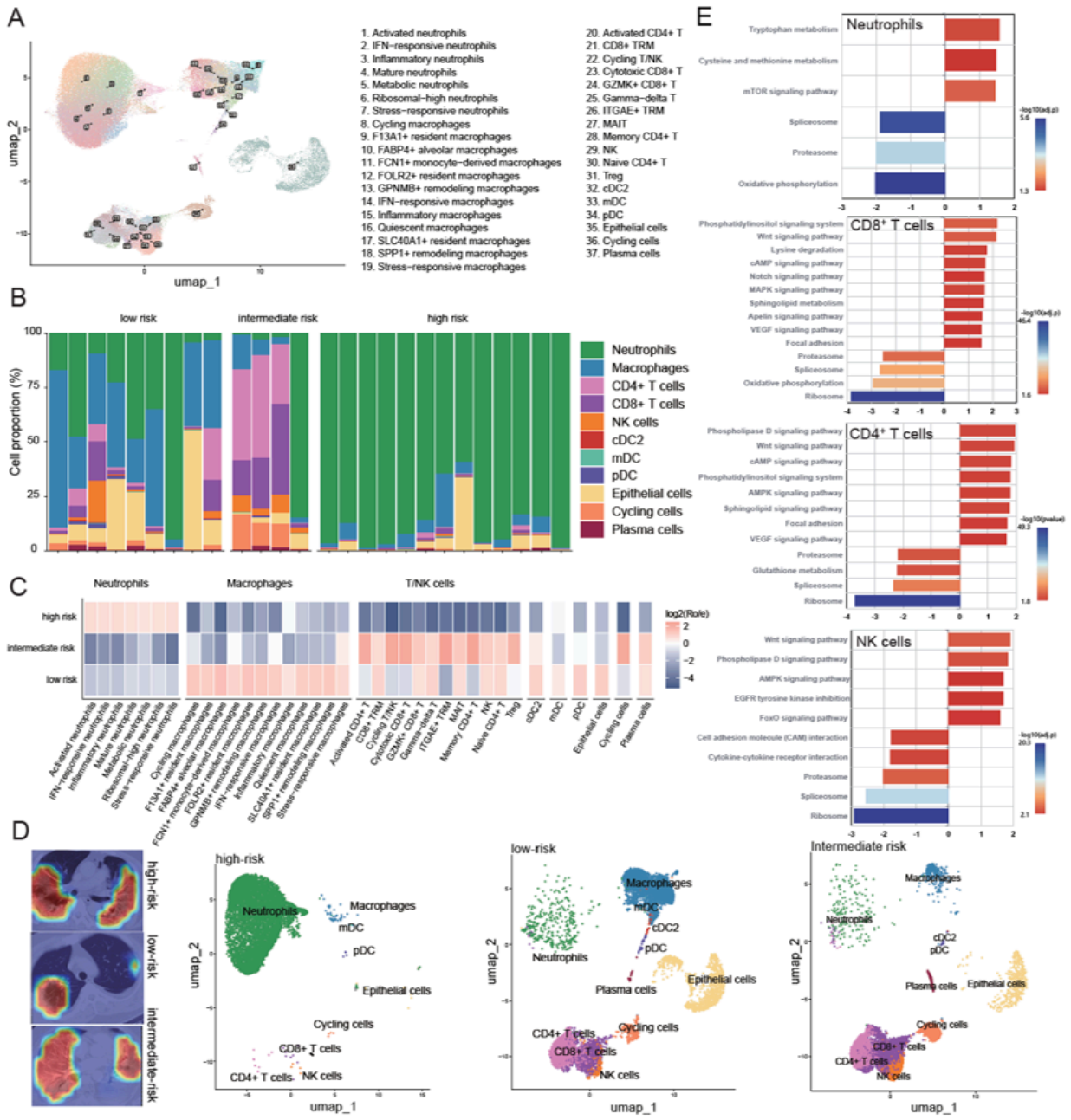



**Figure 6**

**Proteomic differences and biological pathway alterations across risk groups defined by the clinical-imaging and multimodal models.**(**A-C**) Volcano plots showing differential abundant proteins across risk strata defined by the clinical-imaging model, including high- versus low-risk (A), high- versus intermediate-risk (B), and intermediate-

versus low-risk (C) comparisons. Differentially abundant proteins (DAP) were defined by an absolute fold change of ≥1.3 and a false discovery rate (FDR) <0.05. Red and blue points indicate significantly upregulated and downregulated proteins, respectively, whereas gray points indicate nonsignificant proteins. Selected top DAPs are labeled. (**D**) Venn diagrams showing the overlap of DAPs identified from the corresponding pairwise risk-group comparisons derived from the clinical-imaging model and multimodal model. The accompanying percentage bars indicate the proportion of shared and model-specific DAPs in each comparison. (**E**) Heatmap showing the expression patterns of the 25 proteins shared among the overlapping differentially abundant proteins identified in the three pairwise risk-group comparisons in panel D (high versus low risk, high versus intermediate risk, and intermediate versus low risk). Columns represent individual patients, with annotations indicating multimodal-model risk group, clinical-imaging-model risk group, and survival outcome. (**F**) Pathway enrichment analysis of the DAPs identified across the six pairwise risk-group comparisons from the clinical-imaging and multimodal models. Dot size represents statistical significance as −log10 (*P*), and dot color represents the enrichment Z score. Positive Z scores indicate predicted pathway activation, whereas negative Z scores indicate predicted pathway inhibition. Pathways are grouped into major biological categories, including inflammation and immune activation, metabolic and nuclear receptor signaling, extracellular matrix remodeling and organ dysfunction, endothelial injury and vascular dysfunction, and coagulation and thromboinflammation. Abbreviations: HR, high risk; IR, intermediate risk; LR, low risk.




## Figure 7

**Pulmonary immune landscape across multimodal risk groups revealed by BALF single-cell RNA sequencing.** (**A**) Uniform manifold approximation and projection (UMAP) of bronchoalveolar lavage fluid (BALF) single-cell RNA-sequencing data, showing the identified immune and epithelial cell populations and their transcriptionally defined subtypes. Major populations included neutrophils, macrophages, CD4⁺ T cells, CD8⁺ T cells, NK cells, dendritic cells, epithelial cells, cycling cells, and plasma cells. (**B**) Relative proportions of major cell populations across individual BALF samples stratified into low-, intermediate-, and high-risk groups. Each bar represents one patient and shows the percentage composition of the indicated cell populations. (**C**) Relative enrichment of individual cell subtypes across the three risk groups. Colors represent the log2 ratio of observed to expected cell abundance [log2(Ro/e)], with positive values indicating enrichment and negative values indicating depletion within the corresponding risk group. (**D**) Representative chest CT images with model-derived class activation map (CAM) overlays and matched

BALF single-cell transcriptomic profiles across multimodal risk groups. CAMs were superimposed on the original CT images to visualize regions contributing to the deep-learning prediction, with warmer colors indicating greater model activation and cooler colors indicating lower activation. The overlays are shown for visualization and do not represent quantitative lesion segmentation. (**E**) Gene set enrichment analysis comparing the high-risk with the low-risk group within neutrophils, $CD8^+$ T cells, $CD4^+$ T cells, and NK cells. Bars indicate normalized enrichment scores (NES). Positive NES values indicate pathways enriched in the high-risk group, whereas negative NES values indicate pathways enriched in the low-risk group. Bar color represents pathway significance as −log10 (adjusted *P* value). High-risk BALF was characterized by activation of signaling and metabolic programs in T cells and neutrophils, together with suppression of ribosomal, spliceosomal, proteasomal, and oxidative phosphorylation-related programs in several immune cell populations.

## Supplementary Files

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

- supportinginformation.docx