RESEARCH ARTICLE

The role of biomarkers in the prediction and treatment of severe pneumonia

Xiaoqin Huang[†], Min Qing[†], Yu Yin[*]

Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan, China.

Received: February 12, 2026; **accepted:** July 10, 2026.

Severe pneumonia is associated with high morbidity and mortality in critically ill patients. Early identification of patients at risk of adverse outcomes remains essential for timely intervention and individualized treatment. Biomarker-guided decision-making may help improve treatment precision and reduce unnecessary interventions. This study developed an interpretable framework for early risk prediction, key biomarker identification, treatment benefit stratification, and dynamic monitoring in patients with severe pneumonia using real-world intensive care electronic health record data from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. An integrated analytical framework was constructed and Biological Semantic Network Integration International Federation (BioSNI-IF), a joint imputation–prediction model, was proposed. Clinical variables were screened through missingness control and redundancy reduction, and 58 features were retained for modeling. The model incorporated missing-data patterns through a dual-branch architecture and combined joint optimization of imputation and prediction with prior-controllable attention to capture complex nonlinear risk associations. In a cohort of 18,800 patients, BioSNI-IF demonstrated better discrimination and calibration than multiple machine-learning baseline models, while maintaining robust screening performance under specificity-constrained thresholding. Interpretability analyses identified an informative biomarker panel, and stratified analyses suggested heterogeneous treatment effects, indicating that the model might help identify patients more likely to benefit from treatment. Furthermore, early dynamic trajectories of key biomarkers improved risk re-stratification and treatment-response monitoring. This proposed framework provided a practical and interpretable approach for early warning, biomarker discovery, and individualized treatment decision-making in severe pneumonia.

Keywords: severe pneumonia; biomarkers; risk prediction; treatment benefit stratification; dynamic monitoring; precision medicine.

[*]**Corresponding author:** Yu Yin, Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan 637000, China. Email: LeonaVargas145@hotmail.com.

[†]These authors contributed equally to this work.

Introduction

Pneumonia remains a major global health challenge, and lower respiratory tract infections continue to rank among the leading causes of death worldwide [1]. Among affected patients, those who progress to severe pneumonia often develop hypoxemia, overwhelming systemic inflammation, and multiple organ dysfunction, resulting in rapid clinical deterioration and high mortality [2]. In critically ill populations, timely identification of patients at high risk of adverse outcomes is essential for early intervention, rational allocation of medical resources, and optimization of treatment strategies [3]. In recent years, increasing attention has been

directed toward improving prognostic evaluation in severe pneumonia. Conventional assessment tools including vital signs, imaging findings, and established severity scores remain widely used in clinical practice and provide useful support for general risk stratification [4]. Meanwhile, laboratory biomarkers have shown considerable potential because they objectively reflect key pathophysiological processes such as inflammation, coagulation abnormalities, immune dysregulation, and organ injury [5].

Compared with static clinical observations, biomarkers may offer more dynamic and biologically relevant information for early prognosis evaluation and longitudinal disease monitoring. In parallel, machine learning (ML) methods have been increasingly applied to critical care risk prediction with prior studies demonstrating their value in mortality prediction, external validation, reproducibility, and threshold-based clinical application [6-8]. Moreover, the emergence of reporting and appraisal frameworks such as TRIPOD + AI and PROBAST + AI have further promoted transparent model development and more rigorous evaluation of clinical applicability [9, 10]. Explainable artificial intelligence (XAI) has also become an important focus as clinically useful prediction models should not only achieve high performance but also provide interpretable and actionable explanations for physicians [11, 12]. Despite these advances, several important challenges remain unresolved. Severe pneumonia is highly heterogeneous with respect to comorbidities, pathogens, inflammatory responses, and host immune status, meaning that patients with seemingly similar presentations may experience markedly different prognoses and treatment responses [13]. In addition, current tools often perform inadequately during the ultra-early stage of admission, when clinical information is incomplete and missing data are common, thereby limiting stable prediction and potentially contributing to delayed treatment or inefficient resource use [14, 15]. Further, most existing studies focus primarily on outcome prediction, whereas the connection between early risk assessment, biomarker identification, treatment-benefit stratification, and dynamic therapeutic monitoring has not been sufficiently explored. Although corticosteroid and antimicrobial strategies have been investigated in severe community-acquired pneumonia and intensive care unit (ICU) settings, identifying which patients are most likely to benefit from specific treatments remains a major clinical challenge [16–19].

This study systematically investigated the role of biomarkers in the prediction and treatment of severe pneumonia to develop an early risk prediction model for adverse outcomes using intensive care data, identify stable and interpretable biomarkers that could form a clinically practical biomarker panel, and evaluate whether these biomarkers could support treatment-benefit stratification and dynamic monitoring during disease progression. Medical Information Mart for Intensive Care IV (MIMIC-IV) database was employed to provide a reproducible foundation for quantitative modeling, facilitate transparent methodological reporting, and support future external validation and cross-institutional generalizability assessment [20]. In addition, by focusing on clinical information available during the first 24 h after admission, this study aligned the prediction task with real-world early clinical decision-making and reduced the risk of information leakage [21]. An interpretable analytical framework that integrated early clinical information, missing-data-aware modeling, biomarker prioritization, and treatment-related stratified analysis was then constructed. In addition to risk prediction, the framework provided both global and individual-level interpretation of biomarker contributions and assessed whether dynamic biomarker changes could improve follow-up risk re-stratification and treatment-response evaluation. This study proposed a more precise and practical decision-making framework for severe pneumonia, which might help improve early warning, support individualized treatment selection, and promote

the clinical translation of biomarker-based precision management in critically ill patients.

Materials and methods

Data source

This retrospective cohort study utilized the MIMIC-IV database (Beth Israel Deaconess Medical Center, Boston, MA, USA) encompassing admission records from 2008 to 2022, which was a publicly available, de-identified electronic health record database, distributing *via* PhysioNet (https://physionet.org/). All data collection procedures and usage were approved by the Institutional Review Board of Beth Israel Deaconess Medical Center, Boston, MA, USA. Adult patients with pneumonia were identified from initial ICU-related records according to ICD-10 codes J12–J18 and ICD-9 codes 480–486. A total of 28,640 records were initially screened. After excluding 1,120 records from patients younger than 18 years, 6,980 records related to repeated ICU admissions or inter-unit transfers, and 1,740 admissions with a length of stay shorter than 24 h, the final analysis cohort included 18,800 patients with 11,240 males and 7,560 females, aged from 18 to 91 years with a median age of 67 years. The primary reason for hospital admission was pneumonia or suspected lower respiratory tract infection. The main indications for ICU admission were acute hypoxemic respiratory failure in 9,860 patients (52.4%), hemodynamic instability or septic shock in 4,120 patients (21.9%), and the need for advanced respiratory or circulatory support in 4,820 patients (25.6%). Admission time was defined as the index time. Predictor variables were extracted from the first 24 h after admission, and the prediction window was defined as the period from 24 h to 7 days after admission. During this period, 5,080 patients experienced the composite endpoint, corresponding to an event rate of 27.02%, whereas 13,720 patients did not experience endpoint events. The composite endpoint of early deterioration or progression to severe disease was defined as the occurrence of at least one of the events during the prediction window [22], which included transfer from a general ward to the ICU, initiation of vasopressor therapy, all-cause mortality, or initiation of invasive mechanical ventilation. Among patients who met the composite endpoint, 1,520 experienced ward-to-ICU transfer, 2,450 required vasopressor therapy, 720 died, and 720 received invasive mechanical ventilation. However, these component events were not mutually exclusive. The sum of the individual event counts exceeded the total number of patients who reached the composite endpoint. To reduce bias caused by repeated admissions, only the first hospitalization meeting the inclusion criteria were retained.

Feature processing

This study focused on clinical data available in the early post-admission period (0 – 24 h). Candidate features were pooled according to clinical logic, covering demographics, comorbidities, vital signs, respiratory support, laboratory indices, and severity scoring systems, which covered 62 items across seven categories including demographics and admission information of age, sex, body mass index (BMI), smoking history, source of admission, and time from hospital admission to ICU admission; comorbidities/risk factors of COPD/asthma, coronary heart disease/heart failure, diabetes mellitus, hypertension, chronic kidney disease, chronic liver disease, malignancy, immunosuppression, history of stroke, and history of pneumonia/recurrent infections; vital signs and respiratory support of heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate (RR), temperature, peripheral oxygen saturation (SpO_2), fraction of inspired oxygen (FiO_2), oxygen therapy modality, 24-hour urine output, Glasgow Coma Scale (GCS), vasopressor use, and invasive mechanical ventilation; complete blood count (CBC); biochemistry and organ function of albumin (ALB), total bilirubin (TBil), alanine aminotransferase (ALT), aspartate aminotransferase (AST), creatinine (Cr), blood urea nitrogen (BUN), sodium (Na), potassium (K),

glucose (Glu), lactate dehydrogenase (LDH), lactate, and glomerular filtration rate (eGFR); coagulation and inflammation of C-reactive protein (CRP), procalcitonin (PCT), interleukin-6 (IL-6), ferritin, D-dimer, fibrinogen, international normalized ratio (INR), activated partial thromboplastin time (APTT), erythrocyte sedimentation rate (ESR); severity scores of Sequential Organ Failure Assessment (SOFA) score (https://www.mdcalc.com/calc/691/sequential-organ-failure-assessment-sofa-score) and Acute Physiology and Chronic Health Evaluation II (APACHE II) score (https://www.mdcalc.com/calc/1868/apache-ii-score). To minimize the impact of unit inconsistency and entry errors arising from multi-source records on modeling, raw variables were standardized and cleaned [23]. Units for the same indicator across different recording sources were mapped and unified. Obvious irrational values and extreme outliers were subjected to truncation or winsorization to suppress the influence of noise on model estimation and splitting rules. Binary variables were encoded as 0/1, multi-class variables were processed using dummy encoding, and continuous variables underwent z-score standardization. To prevent data leakage, the mean and standard deviation (SD) were estimated solely by using the training set, and these statistics were subsequently applied to the validation and test sets. Based on the raw variables, 16 derived biomarkers were constructed to enhance the characterization of critical pathophysiological processes including NLR (NEU/LYM), PLR (PLT/LYM), SII (PLT × NEU/LYM), CRP/ALB, BUN/ALB, Lactate/ALB, Shock Index (HR/SBP), S/F ratio (SpO_2/FiO_2), eGFR (CKD-EPI), and AST/ALT for inflammatory intensity, immune imbalance, perfusion disorders, and oxygenation impairment. To ensure temporal consistency, all derived indices were calculated using observations solely within the 0 – 24 h window. In cases where the denominator was missing or extremely small leading to unstable ratios, the derived indicator was set to missing, and a missingness mask was retained to avoid artificially introducing irrational extreme ratios. Given the variability in testing strategies and the presence of missing-not-at-random data in real-world ICU settings, missingness was evaluated for all candidate variables. To improve reproducibility and model applicability across clinical settings, a threshold of 30% was applied at the variable level. The highest missing rates were observed for interleukin-6 (58.0%), ferritin (46.0%), and procalcitonin (32.0%), all of which exceeded the predefined threshold of 30% and were therefore excluded from the candidate feature pool. Several additional variables showed moderate missingness including D-dimer (28.0%), NT-proBNP (25.0%), P/F ratio (22.0%), troponin (21.0%), and fibrinogen (20.0%), whereas most routine vital signs, biochemical indices, and complete blood count variables had relatively low missingness. Patients were not excluded because of incomplete data. Instead, missing values in retained variables were handled using a reproducible imputation strategy combined with missingness masks.

BioSNI-IF model architecture

BioSNI-IF model was designed to achieve robust prediction and intrinsic interpretable analysis of risks regarding early deterioration or progression to severe pneumonia. The proposed model combined missing patterns, feature dependency structures, and outcome predictions within a unified framework (Figure 1). The model leveraged clinical data embedded in the presence or absence of test results by explicitly encoding missingness masks. Furthermore, by constructing controllable priors based on statistical correlation structures and introducing a controlled prior feature attention (CPFA) mechanism into the multi-head self-attention module, this prior information was integrated into the representation learning process as soft constraints. This approach established a learnable trade-off between interpretable stability and data-driven expressive power. Following feature construction and time-window aggregation, a structured feature vector was generated for each patient to form the feature matrix X. Simultaneously, a missingness mask matrix M was constructed to incorporate missing

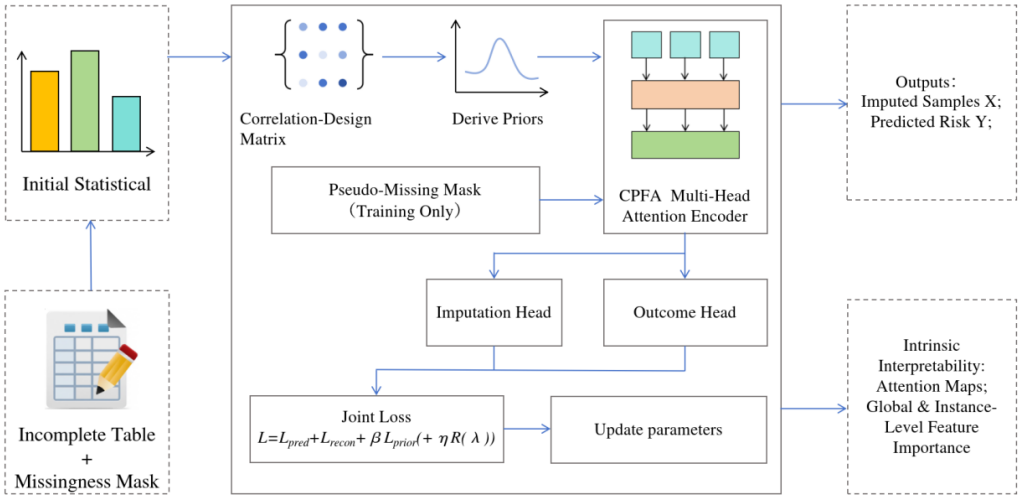

Figure 1. Flowchart of the proposed BioSNI-IF framework.

patterns as part of the model input as below.

$$X \in R^{(n\times d)}, \quad M \in \{0,1\}^{(n\times d)} \tag{1}$$

where n was the sample size. d was the number of features. The model treated each variable as a feature token. Continuous variables utilized a joint encoding of "value + missing indicator" to simultaneously preserve numerical scale and availability information. Categorical variables employed learnable embeddings with a dedicated "missing token" to prevent missing data from being potentially misinterpreted as a valid category. This design enabled the model to learn latent associations between clinical decisions to test or not test specific indicators and disease severity, resource allocation, or physician decision-making, thereby enhancing robustness in real-world scenarios with missing data. To introduce a stable correlation structure early in the training phase and serve as an iterative starting point, lightweight initialization imputation was performed on missing entries in the training set to obtain an initial completed matrix as follows.

$$X^{(0)} = M \odot X + (1 - M) \odot X_{init} \tag{2}$$

where $X^{(0)}$ was the initialized completion matrix used to estimate statistical priors and initiate iterative updates, which served solely to provide a computable statistical structure for constructing priors, not to replace the modeling of missing patterns. X_{int} was the matrix of initial imputation values using medians/quantiles for continuous variables and mode for categorical variables within the training set [24]. The model explicitly used (X, M) as input during both training and inference. Based on the correlation structure estimated from X^g (the completion result of the g-th round), two types of prior distributions were constructed including the feature-feature prior P_f , used to constrain dependencies during imputation, and the feature-outcome prior $P_{(y)}$, used to guide attention during prediction. These priors were input into the model as normalized probability distributions, serving as soft constraint targets for attention learning to enhance structural stability and interpretability consistency. BioSNI-IF employed a self-attention-based encoder to learn interaction relationships between biomarkers. For the h-th attention head, the attention weights were calculated using standard scaled dot-product attention as follows.

$$A^{(h)} = \mathrm{softmax}\left(\frac{Q^{(h)}\left(K^{(h)}\right)^{\top}}{\sqrt{d_k}}\right) \tag{3}$$

where $\mathrm{A}^{(h)}$ was the attention weight matrix for the h-th head, interpretable as the distribution of feature-to-feature dependency intensity. $\mathrm{Q}^{(h)}$ and $\mathrm{K}^{(h)}$ were the query and key representations for the h-th head, obtained *via* linear mapping of token representations. d_k was the dimension of K and Q, used for scaling to stabilize training. To inject statistical priors into the attention mechanism and achieve a controllable trade-off, the model introduced a non-negative intensity parameter λ_h for each attention head. This value was derived from learnable parameters *via* non-linear mapping and regulated the degree to which the head emphasized consistency with the prior. When λ_h was large, the attention head tended to align with the prior, enhancing structural stability and interpretability consistency. When λ_h was small, the head was permitted to deviate from the prior to capture more complex non-linear interactions [25]. At the architectural level, the raw input features were first transformed by a value projection layer and then combined with feature embeddings and positional encodings to form the encoder input tokens. These tokens were processed by the multi-head feature attention module, in which the prior-derived λ_h controlled the extent to which each head followed prior structural information. The attended representations were subsequently passed through an Add&Norm operation and then summarized by a weight aggregation module guided by feature importance scores to generate a compact patient-level representation. In parallel, a feature-bypass pathway directly propagated original input-feature information to the downstream fusion stage, allowing salient low-level clinical signals to be preserved. The fused representation was then refined by a second Add&Norm operation and a feed-forward block followed by final normalization before being forwarded to the classifier. Building upon the shared encoder output, the model adopted a dual-task design. The imputation head output conditional predictions for missing entries to generate the completion matrix X, while the prediction head output the outcome risk probability $\hat{y}$ based on the aggregated patient representation. This joint training ensured that the completion process was no longer a generic imputation independent of the endpoint, but a task-oriented imputation focused on early deterioration or progression to severe disease, which resulted in more stable predictions and more rational feature dependency structures even with highly incomplete clinical data.

Training and evaluation

To enhance generalization capabilities regarding missing mechanisms during the training phase, a portion of the observed entries was randomly masked as "pseudo-missing" to require the model to reconstruct these masked values, thereby introducing a self-supervised reconstruction signal. The model was optimized jointly *via* three loss functions including outcome prediction loss $\mathcal{L}_{\mathrm{pred}}$, imputation reconstruction loss $\mathcal{L}_{\mathrm{recon}}$, and prior constraint loss $\mathcal{L}_{\mathrm{prior}}$ as follows.

$$\mathcal{L} = \mathcal{L}_{\mathrm{pred}} + \alpha\mathcal{L}_{\mathrm{recon}} + \beta\mathcal{L}_{\mathrm{prior}} \quad (4)$$

The outcome prediction loss was implemented using weighted binary cross-entropy (BCE) as below [26].

$$\mathcal{L}_{\mathrm{pred}} = -\frac{1}{n}\sum_{i=1}^{n}[w_1 y_i \log(\hat{y}_i) + w_0(1-y_i)\log(1-\hat{y}_i)] \quad (5)$$

where y_i in {0, 1} was the true outcome for the $i-th sample with 1 indicating early deterioration/progression to severe disease and 0 indicating non-occurrence. $\hat{y}_i$ in (0, 1) was the predicted probability. w_1 and w_0 were positive and negative class weights used to mitigate class imbalance, determined by the inverse of class frequency or hyperparameter tuning on the validation set. The imputation reconstruction loss calculated the error at pseudo-missing positions as below [27].

$$\mathcal{L}_{\mathrm{recon}} = \frac{1}{|\Omega_{\mathrm{pm}}|}\sum_{(i,j)\in\Omega_{\mathrm{pm}}} \ell_{\mathrm{rec}}\left(x_{ij}, \hat{x}_{ij}\right) \quad (6)$$

where Ω_{pm} was the set of pseudo-missing positions randomly masked from originally

observed data. Ω_{pm} was the count of pseudo-missing entries used for averaging. x_{ij} was the true observed value at the masked position. $\hat{x}_{ij}$ was the reconstruction output (imputation head output) for that position. ℓ_{rec} was the single-entry reconstruction error function (squared error for continuous variables, cross-entropy for categorical variables), calculated separately by variable type and aggregated [28]. The prior constraint loss measured the consistency trend between the attention distribution and the prior distribution as follows [29].

$$\mathcal{L}_{\mathrm{prior}} = \sum_{h=1}^{H} \lambda_h \left[\frac{1}{d} \sum_{j=1}^{d} \mathrm{KL}\left(\mathrm{A}_{j,:}^{(h)} \parallel \mathrm{P}_f(j,:) \right) + \mathrm{KL}\left(\mathrm{a}^{(h)} \parallel \mathrm{P}^{(y)} \right) \right] \quad (7)$$

where h was the number of attention heads. λ_h was the prior intensity parameter (learnable) for the λ_h head, determining the influence of the prior constraint. d was the number of feature tokens (real features). $\mathrm{A}_{j,:}^{(h)}$ was the attention distribution of the j-th feature token in the h-th head. $\mathrm{P}_f(j,:)$ was the feature-feature prior distribution indicating which features should be prioritized when completing target feature j. $\mathrm{a}^{(\mathrm{h})}$ was the feature attention distribution used for prediction (typically the attention of the CLS token to feature tokens). P(y) was the feature-outcome prior distribution indicating features relevant to the outcome. KL(p // q) denoted the Kullback-Leibler divergence, where a smaller value indicated greater consistency between distributions p and q [30]. Following feature engineering, missing data handling, and redundancy reduction, 58 features were ultimately retained for modeling. Given the clinical necessity to control false alarms to prevent unnecessary escalation of monitoring and resource consumption, a classification threshold was determined based on a constraint of specificity ≥ 0.90 on the validation set. Specifically, potential thresholds were iterated to identify a candidate set satisfying specificity ≥ 0.90, selecting the threshold that maximized sensitivity under this constraint as the final operating point. This threshold remained fixed for the test set to generate the confusion matrix and calculate sensitivity (Sens), specificity (Spec), positive predictive value (PPV), and negative predictive value (NPV), enabling a comparison of model error structures and screening utility under a unified operating point [31]. Model performance was evaluated across three dimensions including discriminative ability, probabilistic reliability, and clinical utility [32]. The area under the receiver operating characteristic curve (AUROC) was used to measure overall capacity to distinguish and rank "events *vs*. non-events." The area under the precision-recall curve (AUPRC) was employed to focus on the quality of positive identification and prediction precision in the context of relatively scarce events [33]. The Brier score was used to quantify probabilistic prediction error, reflecting the consistency between predicted probabilities and actual outcomes with lower scores indicating better performance [34]. Decision curve analysis (DCA) was utilized to assess the net benefit of intervention strategies across different risk thresholds [35]. Under the fixed threshold, Sens (ability to capture true events, reflecting risk of missed diagnosis), Spec (ability to exclude non-events, reflecting burden of false alarms), PPV (proportion of true events among predicted high-risk cases, reflecting resource hit rate), and NPV (proportion of true non-events among predicted low-risk cases, reflecting reliability of exclusion) were reported. A confusion matrix was used to visually present the composition of false positives and false negatives [36].

Treatment analysis

Following the output of multi-head attention weights and their intensity parameters, a feature dependency structure, feature dependency network, was formed for both global and individual-level explanation. At the global level, the attention dependency structure could be aggregated across samples and resampling iterations to obtain a stable ranking of variable importance. At the individual level, personalized key indicator contributions could be derived based on single-sample attention outputs, achieving "prediction-explanation" consistency [37]. Based on these intrinsic interpretation results, this study screened key biomarkers to

form a core panel by synthesizing importance stability, validation set consistency, missing rates, and clinical availability, which were subsequently used for heterogeneity of treatment effect stratification and dynamic therapeutic monitoring analysis. To validate against traditional interpretation frameworks, SHAP values were calculated on baseline tree models or simplified models to test consistency with the intrinsic importance ranking, thereby enhancing the credibility and reproducibility of the findings. To evaluate heterogeneity of treatment effect (HTE), treatment-benefit stratification was performed based on the ten key biomarkers. The binary adverse outcome was used as the endpoint, and treatment status was defined as the intervention variable. Individual risk representations were constructed from the biomarker panel, and treatment effect was estimated using a model incorporating a treatment-by-risk interaction term. For each patient, predicted risks were generated under treated and untreated counterfactual scenarios, and the difference between these two risks was defined as the individual-level absolute risk reduction (ARR), which served as the benefit score. Patients were subsequently stratified into low/no-benefit, moderate-benefit, and high-benefit groups according to the quantiles of the benefit score. To reduce confounding bias inherent in real-world treatment allocation, inverse probability of treatment weighting (IPTW) was applied. Within each benefit stratum, weighted ARR and the corresponding number needed to treat (NNT) were estimated, and interaction testing was used to assess treatment-effect heterogeneity. At the policy level, four treatment strategies were compared including treat-none, treat-all, treat-high-benefit, and treat-high-risk, and their clinical value was evaluated using decision curve analysis across different treatment thresholds. To assess the value of dynamic biomarker monitoring, repeated measurements within 0 – 72 h after admission were used to construct time-series features for the key biomarker panel. Dynamic change features (Δ) were defined across prespecified time windows using admission time as the reference. For clinical interpretability, longitudinal analyses focused on lactate, neutrophil-to-lymphocyte ratio (NLR), and albumin. Trend tests with group-by-time interaction terms were used to examine whether temporal changes differed according to outcome status. In addition, two discriminative models were developed for comparison, which were a baseline model using admission biomarker levels only and a dynamic-enhanced model additionally incorporating Δ features. Their discriminative performance was compared on the test set using the AUROC to quantify the incremental value of dynamic monitoring.

Results and discussion

Cohort description and feature screening

In the feature correlation analysis, retention or merging decisions were made based on clinical interpretability, data availability, and missingness for highly correlated feature pairs, while a selection or grouped control strategy was adopted for features with conceptual overlaps such as oxygenation-related indicators and ratio-derived biomarkers to avoid repetitive representation of the same pathophysiological dimension. By using a threshold of $|\rho| \geq 0.85$, 16 redundant features were identified and removed, while 58 features were retained for subsequent modeling and interpretability analysis. Representative variables being excluded included lymphopenia_flag, creatinine, hypoxemia_flag, low_alb_flag, AST/ALT, lactate/albumin, CRP/albumin, BUN/albumin, high_ddimer_flag, systolic blood pressure, thrombocytopenia_flag, BUN/Cr, diastolic blood pressure, ALT, shock index, and bicarbonate. These variables showed strong correlations with clinically related retained features such as lym_abs, eGFR, S/F ratio, albumin, AST, lactate, CRP, BUN, D-dimer, mean arterial pressure, platelet count, heart rate, and anion gap, indicating substantial redundancy among threshold-derived, ratio-based, and overlapping physiological indicators. The final retained feature pool covered multiple clinically relevant

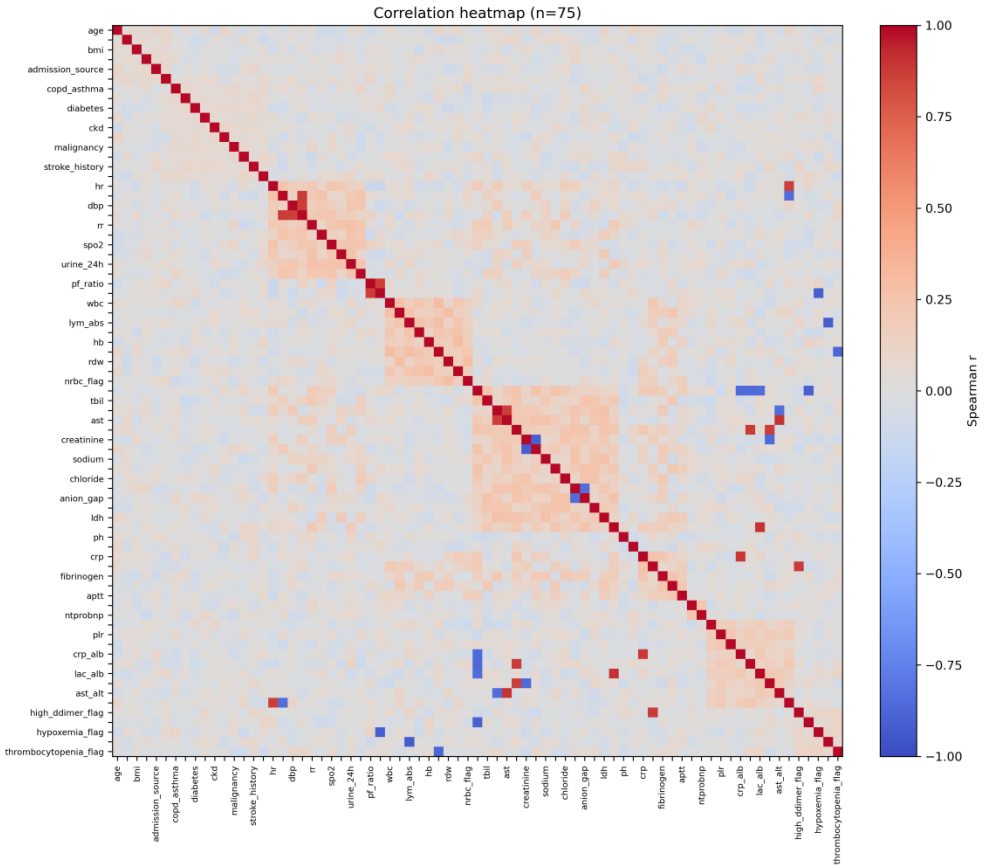

Figure 2. Heatmap of feature correlation analysis.

domains including demographics and admission characteristics, comorbidities, vital signs, oxygenation, complete blood count, biochemistry, metabolic and perfusion markers, blood gas variables, coagulation and inflammation markers, and derived composite indices. Specifically, the retained variables included demographic and admission features such as age, sex, body mass index, smoking status, admission source, and pre-ICU length of stay; major comorbidities such as COPD/asthma, coronary heart disease or heart failure, diabetes, hypertension, chronic kidney disease, chronic liver disease, malignancy, immunosuppression, stroke history, and chronic infection; physiological variables including heart rate, mean arterial pressure, respiratory rate, temperature, SpO_2, FiO_2, urine output, and Glasgow Coma Scale; laboratory variables including white blood cell count, neutrophil percentage, absolute lymphocyte count, monocyte count, hemoglobin, platelet count, RDW, MPV, albumin, total bilirubin, AST, BUN, eGFR, sodium, potassium, chloride, anion gap, glucose, LDH, lactate, pH, PaCO2, CRP, D-dimer, INR, APTT, and fibrinogen; and composite biomarker indices including NLR, PLR, and SII (Figure 2). Collectively, these retained features comprehensively characterized early inflammatory burden, immune imbalance, oxygenation impairment, perfusion and metabolic disturbance, and organ dysfunction in severe pneumonia, thereby providing a structured basis for predictive modeling and biomarker mining.

Predictive performance, calibration, and clinical utility

The results of comparison of AUROC, AUPRC, and Brier scores showed that, overall, all models demonstrated robust discriminative

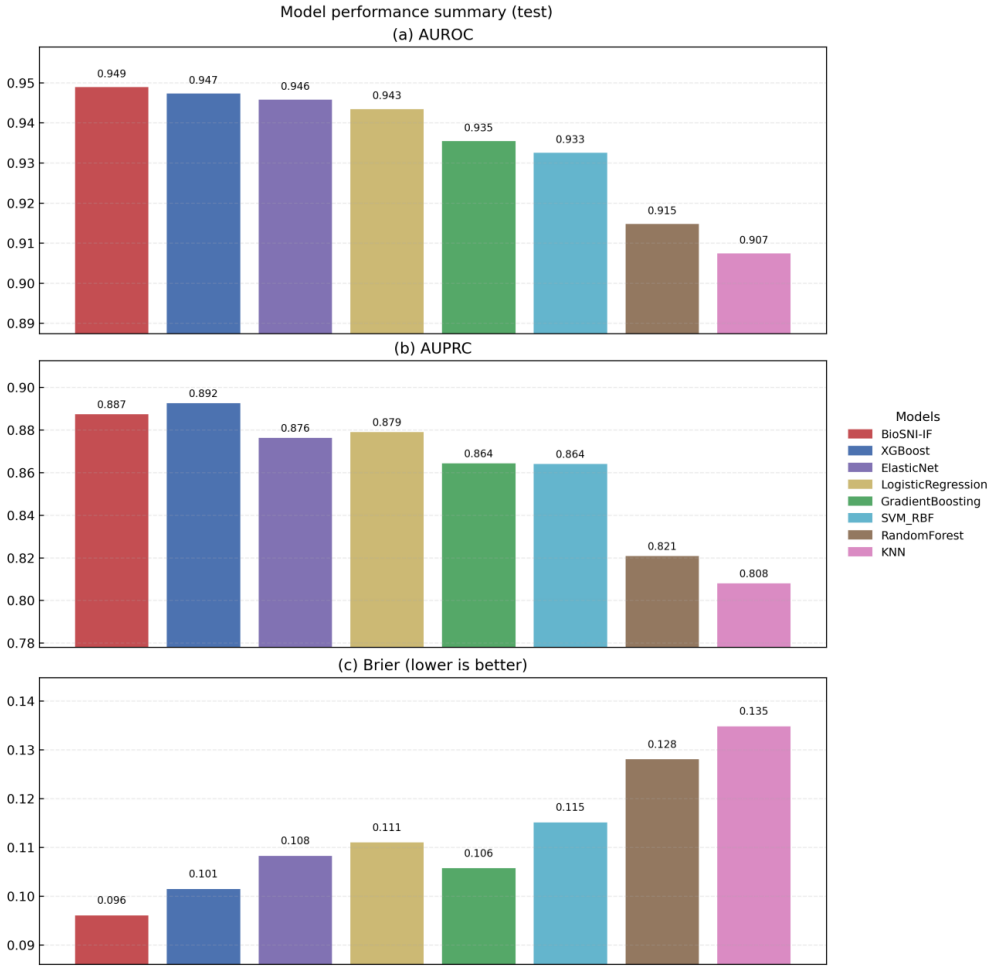

Figure 3. Comparison of model AUROC, AUPRC, and Brier score.

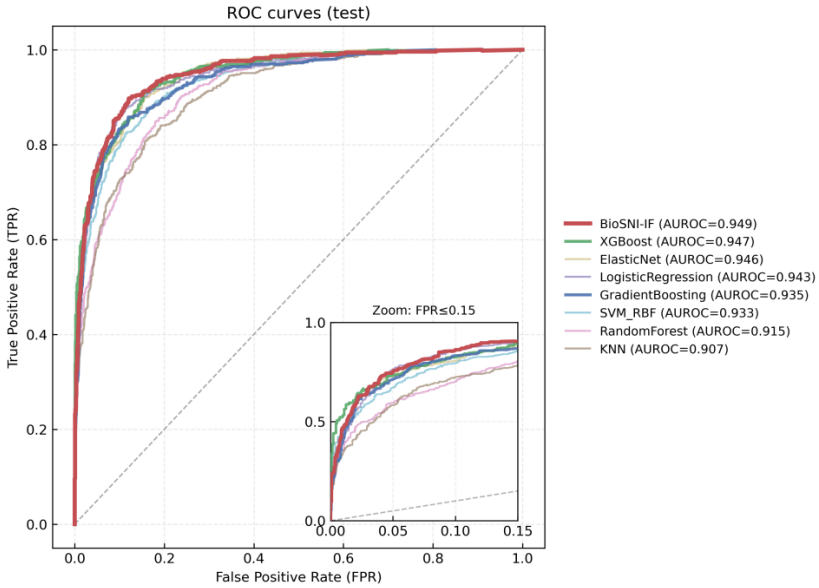

Figure 4. Comparison of ROC curves across models.

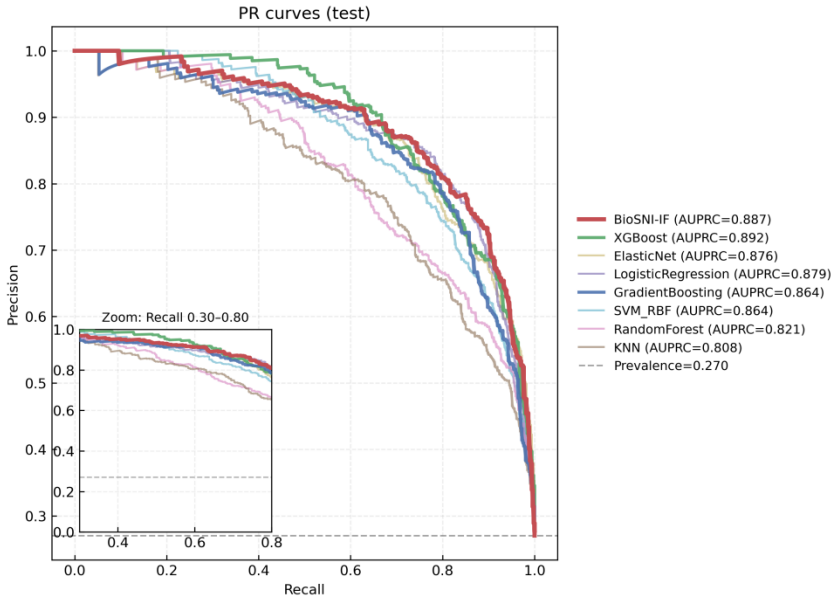

Figure 5. Comparison of precision–recall (PR) curves across models.

performance with BioSNI-IF achieving the best comprehensive performance (Figure 3). The ROC curves for BioSNI-IF and XGBoost generally laid above those of other models, indicating superior discriminative ability across various threshold ranges. Notably, in the low false-positive rate region, both models maintained higher true-positive rates, aligning with the clinical requirement to control false alarms (Figure 4). Given the test set event rate of approximately 0.27, the PR curves for all models were significantly above the baseline, confirming that the models effectively learned outcome-relevant information. In the PR space, BioSNI-IF and XGBoost maintained high precision across a wide recall range, suggesting robust identification of positive samples (Figure 5), which made them particularly suitable for clinical prediction tasks where the event rate was not extremely low, but class imbalance persisted. Based on the validation set strategy (Specificity ≥ 0.90), the classification performance of each model on the test set showed that BioSNI-IF achieved high sensitivity of 0.870 and negative predictive value of 0.949, which implied that it minimized missed detections while controlling false alarms, providing reliable support for "excluding event risk". XGBoost exhibited the highest specificity of 0.921 and a relatively high PPV of 0.788 but lower sensitivity of 0.795, reflecting a more "conservative" strategy for positive classification. Logistic regression maintained competitive sensitivity of 0.841 and NPV of 0.939 at this threshold, indicating that linear models could still offer strong baseline performance. KNN demonstrated the lowest sensitivity of 0.713, suggesting a more pronounced risk of missed diagnosis under the same specificity constraint (Figure 6). The visualization of the confusion matrices further revealed the misclassification patterns of different models. While some models increased false negatives (FN) to improve specificity, BioSNI-IF tended to achieve a relative balance between FN and false positives (FP), thereby forming operating characteristics more suitable for clinical screening and early warning. In clinical applications, model output probabilities were frequently used for risk communication, threshold-based decision-making, and resource allocation, therefore, calibration ability was as critical as discrimination. The calibration curves for each model demonstrated that the BioSNI-IF curve most closely approximated the ideal diagonal, indicating a high consistency between predicted probabilities and actual event rates (Figure 7a).

Figure 6. Confusion matrix of model classification.

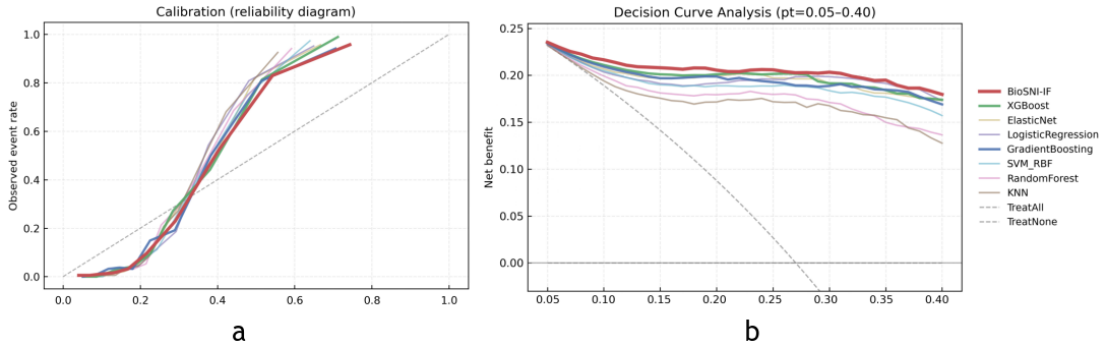

Figure 7. Comparison of calibration curves (**a**) and decision curves (**b**) across models.

This result aligned with its lowest Brier score (0.096). In contrast, other models showed varying degrees of deviation in the medium-to-high risk intervals. Specifically, random forest and KNN yielded higher Brier scores of 0.128 and 0.135, respectively, suggesting larger errors in probability output, which might limit their utility in clinical scenarios requiring precise probabilities rather than binary labels. To evaluate the clinical net benefit of the models across different risk thresholds, decision curve analysis was conducted. Within the given threshold probability range of pt = 0.05 - 0.40, the BioSNI-IF curve was generally higher than or equal to those of other models and superior to the "Treat-none" and "Treat-all" reference strategies (Figure 7b), which indicated that, using BioSNI-IF for intervention decisions within this threshold range yielded a higher net benefit. Combined with its superior calibration performance, the clinical utility advantage of BioSNI-IF was credible. The model was not only discriminative but also probabilistically reliable, making it highly suitable for subsequent stratification of treatment benefits and strategy evaluation.

Model interpretability and key biomarker panel

To characterize the structural coupling relationships between key biomarkers and provide a basis for subsequent biomarker panel refinement and mechanistic explanation, a

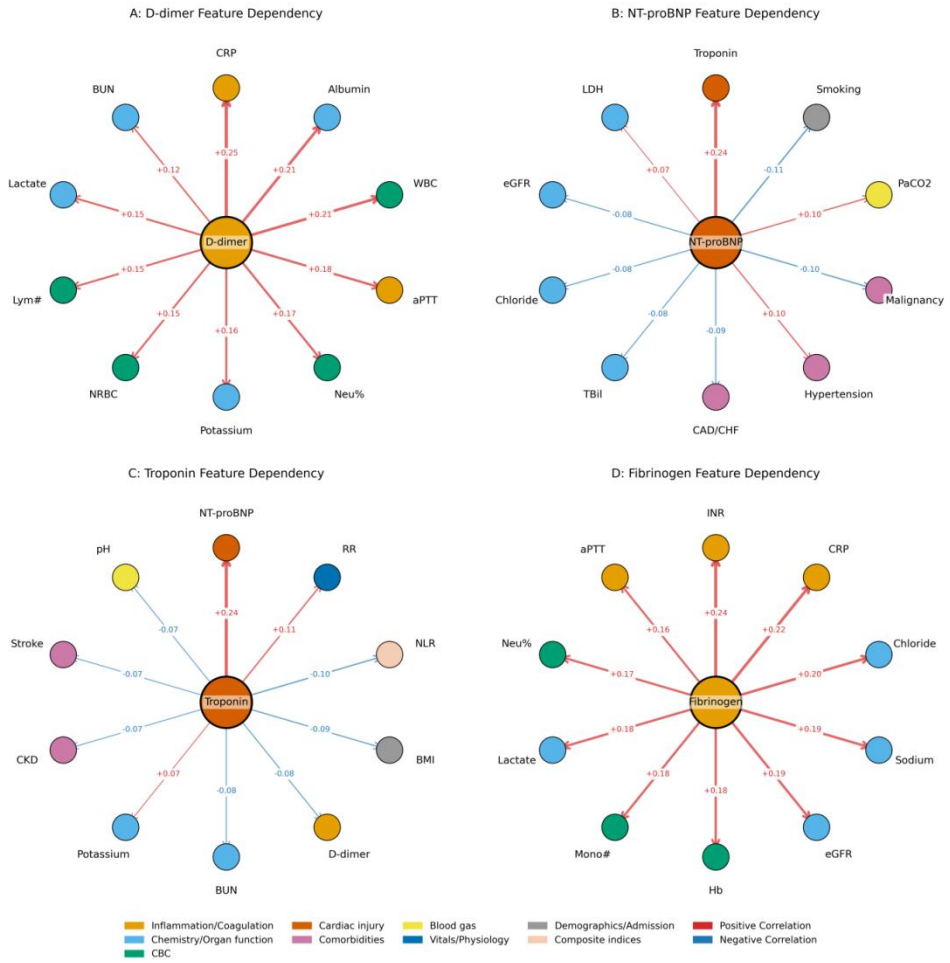

Figure 8. Feature dependency subnetwork.

feature dependency network was constructed using the final feature set employed for modeling. The statistical degree of dependency between any two features was calculated within the same analysis cohort. Given that clinical variable distributions were often skewed and exhibited non-linear monotonic relationships [38], the Spearman rank correlation coefficient (ρ) was adopted as the measure of feature dependency intensity to generate the feature dependency matrix. In the network construction phase, each feature was treated as a node. If any two features satisfied $|\rho| \geq \tau$, where τ was the threshold controlling network sparsity and readability, an edge was created between the nodes. The weight of the edge was defined as $|\rho|$, and the sign of the edge was determined by the sign of ρ. To further enhance visual interpretability, a Top-K sparsification strategy was applied when displaying local dependency structures by centering on a specific key feature that only the top K edges with the highest absolute correlation coefficients linking it to other features were retained. This resulted in a local dependency sub-network centered on the key feature. This network reflected feature-level statistical correlation structures, revealing potential common variation patterns and information redundancy, but did not constitute causal inference. Four local dependency sub-networks centered on D-dimer, NT-proBNP, troponin, and fibrinogen showed variable categories with different node colors, the direction of correlation (positive/negative) with

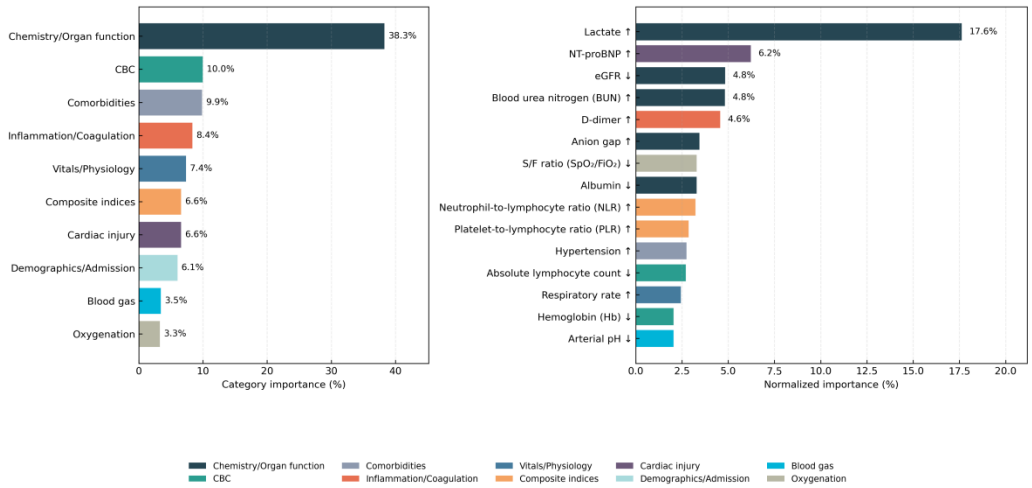

Figure 9. Feature importance ranking.

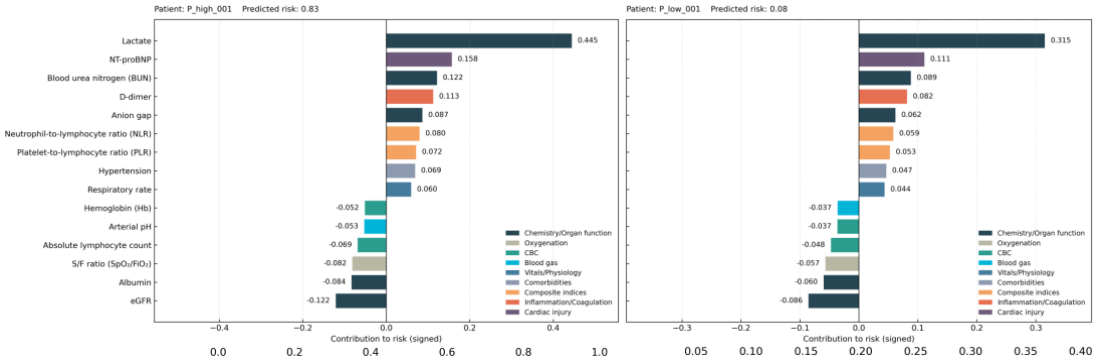

Figure 10. Feature contributions for a high-risk case.

edge colors, the Spearman ρ values as edge labels, while edge thickness was proportional to |ρ| (Figure 8). To enhance the interpretability and clinical translatability of the predictive model, the decision-making rationale of BioSNI-IF was further interpreted. A key biomarker panel was constructed for subsequent treatment benefit stratification and dynamic monitoring. The interpretability analysis included both global and individual explanations. During the inference phase, BioSNI-IF output the contribution of each feature to the predicted risk. To obtain stable conclusions at the global level, feature contributions were converted to absolute values, normalized, and aggregated across the test set to derive a global importance ranking. Simultaneously, features were classified by clinical semantics into categories such as organ function/biochemistry, complete blood count (CBC), inflammation/coagulation, vital signs/physiology, oxygenation, blood gas, and comorbidities. These category-level contributions were aggregated to present the pathological axis composition of the model's decision-making (Figure 9). Category-level aggregation revealed that the model's prediction relied most heavily on organ function/biochemistry-related features (38.3%) followed by CBC (10.0%) and comorbidities (9.9%). Inflammation/coagulation and vital signs/physiology each accounted for 8.4%, composite indices for 6.6%, myocardial injury for 6.0%, demographics/admission information for 4.7%, and blood gas and oxygenation for 3.5%

Table 1. Biomarker selection results.

Rank	Biomarker	Global Importance	Risk Direction	Category
1	Lactate	0.1764	↑ → Risk↑	Perfusion/Metabolism
2	NT-proBNP	0.0624	↑ → Risk↑	Cardiovascular Stress
3	e-GFR	0.0484	↓ → Risk↑	Renal Function
4	BUN	0.0483	↑ → Risk↑	Renal Function/Metabolism
5	D-dimer	0.0457	↑ → Risk↑	Coagulation/Inflammation
6	Anion_gap	0.0346	↑ → Risk↑	Metabolism/Acid-Base
7	S/F_ratio	0.0331	↓ → Risk↑	Oxygenation
8	Albumin	0.0330	↓ → Risk↑	Nutrition/Inflammation
9	NLR	0.0324	↑ → Risk↑	Immunity/Inflammation
10	Lym_abs	0.0272	↓ → Risk↑	Immunity

and 3.7%, respectively. These results suggested that the model's discrimination of adverse outcomes in severe pneumonia did not depend on a single indicator but integrated multiple pathological dimensions, consistent with clinical understanding of the mechanisms driving critical illness evolution. To demonstrate the decision-making rationale of BioSNI-IF at the individual patient level, signed contributions were visualized. The feature contributions for representative high-risk and low-risk cases demonstrated were shown in Figure 10. To form a key biomarker panel suitable for subsequent treatment benefit stratification and dynamic monitoring analysis, candidates were screened based on the reproducible criteria of priority on contribution stability and importance with stable and high contribution indicators to population-level prediction based on the global importance ranking, redundancy reduction by excluding indicators that were highly collinear or semantically repetitive with retained variables through combining results from the correlation analysis, availability and low missingness by giving priority to indicators routinely available in the ICU with controllable missing rates larger than 30% being retained for supplementary analysis but excluded from the core panel, clinical interpretability by selecting indicators mappable to distinct pathological pathways and facilitating the interpretation of their risk implications within a clinical context. Based on these criteria, 10 key biomarkers were ultimately identified to constitute the core panel (Table 1), serving as the unified input variable set for subsequent experiments.

Treatment-benefit stratification and policy comparison

After IPTW adjustment, treatment in the overall population yielded an ARR of approximately 3.6%, corresponding to an NNT of 27.7. Stratified analysis indicated that the high-benefit stratum derived the greatest benefit with ARR approx 8.0% and NNT approx 12.5, whereas the moderate-benefit stratum showed limited benefit with ARR approx 3.5% and NNT approx 28.6. The low/no-benefit stratum exhibited an ARR approaching 0, suggesting uncertain therapeutic benefit. The interaction tests yielded *P* value (interaction) of 0.004, indicating that the treatment effect varied significantly across benefit strata, supporting the ability of biomarker-based stratification to capture THE (Figure 11). Strategic comparison revealed that treating only the high-benefit stratum achieved a better balance between clinical yield and medication burden. Treat-all reduced the expected event rate from 28.9% to 25.3% with ARR approx 3.6% and NNT approx 27.7 but required 100% population coverage. Treat-high-benefit with only 33.3% coverage yielded an expected event rate of 26.2% with ARR approx 2.7% and NNT approx 37.5. Treat-high-risk covered 25.0% of the population, resulting in an expected event rate of 26.6% with ARR approx 2.3% and NNT approx 43.5. These results suggested that, compared to stratification based

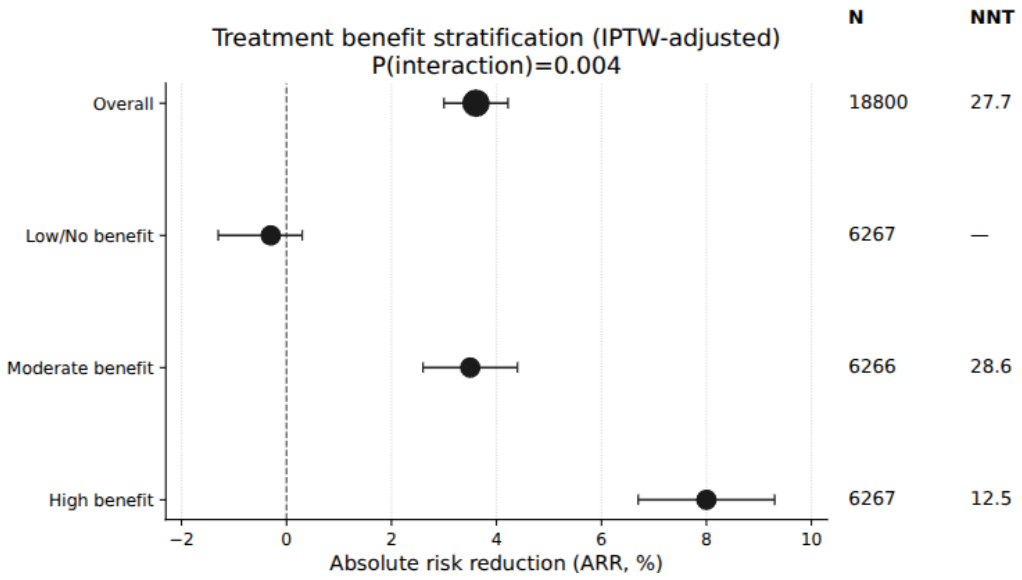

Figure 11. Treatment benefit stratification based on key biomarkers.

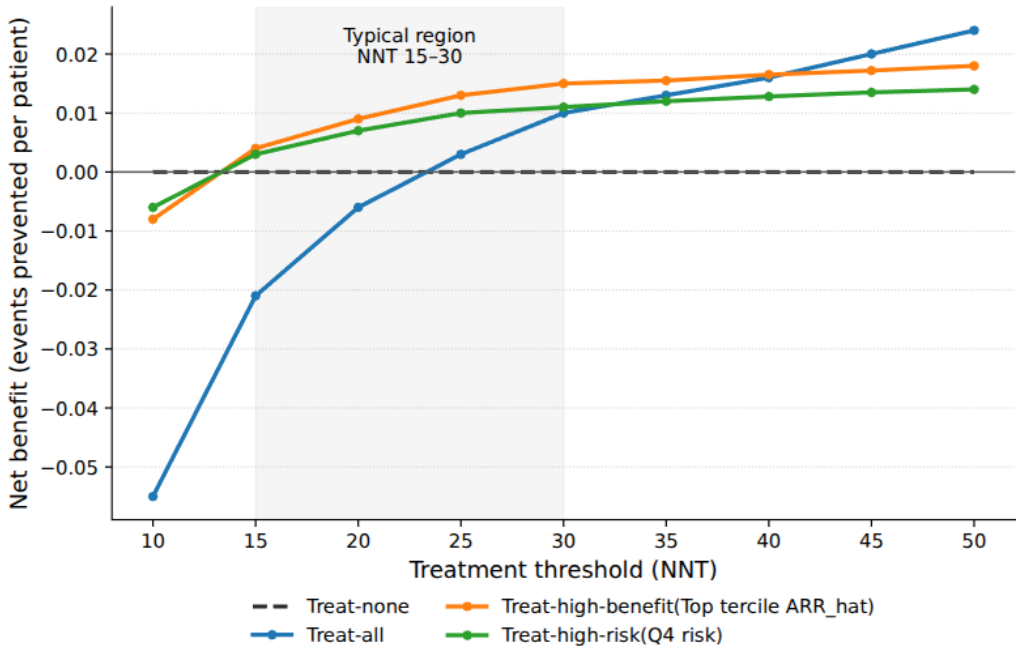

Figure 12. Overall net benefit curves from decision curve analysis (DCA).

solely on risk, the benefit-score-based strategy achieved superior overall yield with lower medication coverage. Within the clinically common threshold range (NNT 15 – 30), the treat-high-benefit strategy demonstrated higher net benefit at most threshold points, indicating an optimal trade-off between reducing unnecessary treatment and maintaining substantial clinical gain. Conversely, treat-all showed potential negative net benefits at stricter thresholds with lower NNT, requiring higher benefit, reflecting the risk of overtreatment under such preferences (Figure 12). Synthesizing the forest plot, strategy table, and net benefit

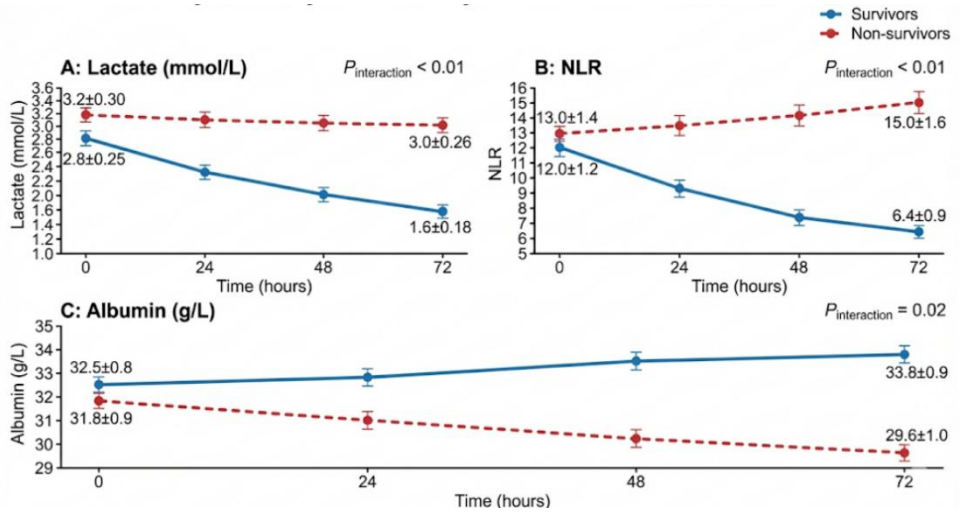

Figure 13. Early dynamic changes in key biomarkers and outcome differences.

curves, this study demonstrated that benefit scores constructed from the ten key biomarkers could effectively identify high-benefit subgroups and hold potential for supporting individualized treatment decisions.

Biomarker dynamic trajectories and therapeutic monitoring

The dynamic-enhanced model outperformed the baseline-only model in outcome discrimination. ROC results indicated that the baseline model achieved an AUC of 0.74, whereas the inclusion of dynamic change features Δ improved the AUC to 0.82, which suggested that early dynamic monitoring provided significant gains for risk identification. These findings implied that relying solely on a single admission measurement might underestimate the heterogeneity of patient outcomes, whereas dynamic features captured disease evolution and treatment response differences, thereby enhancing the model's ability to distinguish adverse outcomes. The 72-hour trajectory analysis of key biomarkers further confirmed the consistency between physiological improvement and outcome differentiation. Lactate showed a sustained downward trend in the survivor group, while the non-survivor group exhibited limited reduction or a plateau, suggesting that insufficient improvement in tissue perfusion and metabolic status was associated with adverse outcomes. NLR decreased significantly over time in survivors but remained high or increased in non-survivors, reflecting the association between persistent inflammatory response/immune imbalance and poor prognosis. Conversely, Albumin gradually recovered in survivors but continued to decline in non-survivors, indicating the protective role of nutritional status and the recovery of inflammation-related protein synthesis (Figure 13). Group-by-time interaction tests for all three markers were statistically significant, indicating systematic differences in dynamic trends between outcome groups, validating their use as quantifiable metrics for therapeutic monitoring. To facilitate clinical interpretation and application, Δ indices were constructed and compared across different time windows. For NLR $\Delta_{0\text{-}48h}$, survivors showed a marked decrease, while non-survivors showed an increasing trend. For Albumin $\Delta_{0\text{-}72h}$, survivors exhibited an increase versus a decrease in non-survivors. All differences were statistically significant. Collectively, these results demonstrated that the early change Δ in key biomarkers could reflect physiological improvement or deterioration shortly after treatment, aligning with changes in subsequent outcome risk. Thus, they served as a crucial basis for therapeutic monitoring and dynamic risk re-stratification.

Conclusion

This study focused on early in-hospital risk prediction for adverse outcomes in patients with severe pneumonia and the evaluation of treatment benefits. A unified analytical framework was developed, which spanned feature engineering, predictive modeling, interpretable biomarker selection, and stratified clinical application. The results demonstrated that a stable set of 58 features covering key domains of vital signs, laboratory measurements, oxygenation status, and baseline comorbidities was identified, providing a reusable structured input foundation for downstream risk modeling and biomarker screening. Based on interpretability-driven selection, a panel of 10 key biomarkers including lactate, NT-proBNP, eGFR, BUN, D-dimer, anion gap, S/F ratio, albumin, NLR, and absolute lymphocyte count was identified, which enabled clinically interpretable characterization of risk heterogeneity along major pathophysiological axes including perfusion–metabolic disturbance, coagulation–inflammatory activation, renal dysfunction, and oxygenation–nutritional status. The proposed BioSNI-IF model explicitly leveraged missingness patterns *via* a dual-branch architecture that jointly optimized imputation and prediction and integrated prior-controllable attention with interpretable output mechanisms to achieve robust modeling of nonlinear risk associations. Beyond providing global feature importance and individualized feature attributions, BioSNI-IF further constructed a feature dependency network to reveal conditional dependency structures among key variables, thereby translating black-box predictions into a more communicable chain of clinical evidence and supporting both biomarker identification and mechanistic interpretation. The empirical performance and clinical utility confirmed that BioSNI-IF achieved strong overall performance on the test set, outperforming multiple machine-learning baselines. Under a specificity-constrained thresholding strategy, the model maintained high screening capability, indicating effective reduction of missed detections while controlling false-positive burden. Treatment stratification analyses further showed that, in the overall population, a treat-all strategy reduced the expected event rate, corresponding to an absolute risk reduction and a number needed to treat. Moreover, model-guided targeted treatment achieved stable net benefits at lower treatment coverage rates. Dynamic monitoring results demonstrated that incorporating early biomarker changes increased the AUC of a dynamically enhanced model, suggesting that biomarker trajectories could support risk re-stratification and treatment-response monitoring, thereby further strengthening clinical applicability.

References

1. GBD 2023 Lower Respiratory Infections and Antimicrobial Resistance Collaborators. 2026. Global burden of lower respiratory infections and aetiologies, 1990-2023: A systematic analysis for the Global Burden of Disease Study 2023. Lancet Infect Dis. 26(4):343-361.
2. GBD 2021 Lower Respiratory Infections and Antimicrobial Resistance Collaborators. 2024. Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990-2021: A systematic analysis from the Global Burden of Disease Study 2021. Lancet Infect Dis. 24(9):974-1002.
3. File TM Jr, Ramirez JA. 2023. Community-acquired pneumonia. N Engl J Med. 389(7):632-641.
4. Martin-Loeches I, Torres A, Nagavci B, Aliberti S, Antonelli M, Bassetti M, *et al*. 2023. ERS/ESICM/ESCMID/ALAT guidelines for the management of severe community-acquired pneumonia. Intensive Care Med. 49(6):615-632.
5. Christ-Crain M, Opal SM. 2010. Clinical review: The role of biomarkers in the diagnosis and management of community-acquired pneumonia. Crit Care. 14(1):203.
6. Cilloniz C, Ward L, Mogensen ML, Pericàs JM, Méndez R, Gabarrús A, *et al*. 2023. Machine-learning model for mortality prediction in patients with community-acquired pneumonia: Development and validation study. Chest. 163(1):77-88.
7. Pan J, Guo T, Kong H, Bu W, Shao M, Geng Z. 2025. Prediction of mortality risk in patients with severe community-acquired pneumonia in the intensive care unit using machine learning. Sci Rep. 15(1):1566.
8. Zhao W, Li X, Gao L, Ai Z, Lu Y, Li J, *et al*. 2025. Machine learning-based model for predicting all-cause mortality in severe pneumonia. BMJ Open Respir Res. 12(1):e001983.
9. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, *et al*. 2024. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 385:e078378.

10. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, *et al*. 2025. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 388:e082505.
11. Bienefeld N, Boss JM, Lüthy R, Brodbeck D, Azzati J, Blaser M, *et al.* 2023. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. npj Digit Med. 6:94.
12. Nicolson A, Bradburn E, Gal Y, Papageorghiou AT, Noble JA. 2025. The human factor in explainable artificial intelligence: clinician variability in trust, reliance, and performance. npj Digit Med. 8(1):658.
13. Martin-Loeches I, Reyes LF, Rodriguez A. 2025. Severe community-acquired pneumonia (sCAP): Advances in management and future directions. Thorax. 80(8):565-575.
14. Dequin PF, Torres A, Bassetti M, Ferrer M, Giamarellos-Bourboulis EJ, Martin-Loeches I, 2026. Severe community-acquired pneumonia: Current concepts and controversies. Intensive Care Med. 52(1):118-130.
15. Jones BE, Ramirez JA, Oren E, Soni NJ, Sullivan LR, Restrepo MI, *et al*. 2026. Diagnosis and management of community-acquired pneumonia. An official American Thoracic Society clinical practice guideline. Am J Respir Crit Care Med. 212(1):24-44.
16. Dequin PF, Meziani F, Quenot JP, Kamel T, Ricard JD, Badie J, *et al*. 2023. Hydrocortisone in severe community-acquired pneumonia. N Engl J Med. 388(21):1931-1941.
17. Smit JM, Van Der Zee PA, Stoof SCM, Van Genderen ME, Snijders D, Boersma WG, *et al*. 2025. Predicting benefit from adjuvant therapy with corticosteroids in community-acquired pneumonia: A data-driven analysis of randomised trials. Lancet Respir Med. 13(3):221-233.
18. See XY, Wang TH, Chang YC, Lo J, Liu W, Choo CYW, *et al*. 2024. Impact of different corticosteroids on severe community-acquired pneumonia: A systematic review and meta-analysis. BMJ Open Respir Res. 11(1):e002141.
19. Duijkers R, Prins HJ, Kross M, Snijders D, van den Berg JWK, Werkman GM, *et al*. 2024. Biomarker guided antibiotic stewardship in community acquired pneumonia: A randomized controlled trial. PLoS One. 19(8):e0307193.
20. Johnson AEW, Bulgarelli L, Shen L, Alvin Gayles, Ayad Shammout, Steven Horng, *et al.* 2023. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 10(1):1.
21. Chan B, Kim B, Schonfeld E, Nageeb G, Pant A, Sjoholm A, *et al*. 2025. Publicly available datasets for artificial intelligence in neurosurgery: A systematic review. J Clin Med. 14(16):5674.
22. Parthasarathi A, Padashetti VC, Padukudru S, Chaya SK, Siddaiah JB, Anand MP. 2022. Association of serum albumin and copeptin with early clinical deterioration and instability in community-acquired pneumonia. Adv Respir Med. 90(4):323-337.
23. Zayed AM, Sarikakis I, Delvaux N. 2026. Automated standardization and harmonization of laboratory units in large-scale clinical data using open-source R functions. Int J Med Inform. 205:106131.
24. Shi J, Hubbard AE, Fong N, Pirracchio R. 2025. Implicit bias in ICU electronic health record data: Measurement frequencies and missing data rates of clinical variables. BMC Med Inform Decis Mak. 25(1):241.
25. Maekaku T, Fujita Y, Peng Y, Watanabe S. 2022. Attention weight smoothing using prior distributions for transformer-based end-to-end ASR. In Interspeech 2022. 2022:1071-1075.
26. Ma H, Bai H, Yan J, Chen Q, Ma Z, Tong S, *et al*. 2025. AI approaches for predicting progression to acute coronary syndrome among stable coronary heart disease patients. npj Cardiovasc Health. 2:59.
27. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. 11:3371-3408.
28. Delong L, Kozak A. 2023. The use of autoencoders for training neural networks with mixed categorical and numerical features. ASTIN Bull. 53(2):213-232.
29. Tachibana H, Uenoyama K, Aihara S. 2018. Efficiently trainable text-to-speech system based on deep convolutional networks with guided attention. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. 2018:4784-4788.
30. Murphy KP. 2022. Probabilistic machine learning: An introduction. Cambridge (MA): MIT Press.
31. Barthélemy NR, Salvadó G, Schindler SE, He Y, Janelidze S, Collij LE, *et al*. 2024. Highly accurate blood test for Alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests. Nat Med. 30(4):1085-1095.
32. Van Calster B, Collins GS, Vickers AJ, Wynants L, Kerr KF, Barreñada L, *et al*. 2025. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: Overview and guidance. Lancet Digit Health. 7(12):100916.
33. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al*. 2010. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology. 21(1):128-138.
34. Saito T, Rehmsmeier M. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 10(3):e0118432.
35. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, *et al*. 2018. Reporting and interpreting decision curve analysis: A guide for investigators. Eur Urol. 74(6):796-804.
36. Vickers AJ, Elkin EB. 2006. Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making. 26(6):565-574.
37. Arik SÖ, Pfister T. 2021. TabNet: Attentive interpretable tabular learning. Proc AAAI Conf Artif Intell. 35(8):6679-6687.
38. Steyerberg EW. 2008. Applications of prediction models. In clinical prediction models: A practical approach to development, validation, and updating. New York: Springer. 11-31.