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.

Integrating Inflammatory and Endothelial Biomarkers in Machine Learning Models to Predict Sepsis Development in Critically Ill Patients: A Combined Retrospective and Prospective Multicenter Study

Ioannis Papapanagiotou
Internet Business Hellas
Alice G. Vassiliou
alvass75@gmail.com

National & Kapodistrian University of Athens
Charikleia S. Vrettou
National & Kapodistrian University of Athens
Maria Theodorakopoulou
General Hospital Attiki KAT
Zafeiria Mastora
General Hospital of Thoracic Diseases "Sotiria"
Apostolos Karalis
Internet Business Hellas
Vassiliki Giannopoulou
National & Kapodistrian University of Athens
Olga Kampouropoulou
National & Kapodistrian University of Athens
Ioannis Charcharos
Internet Business Hellas
Georgios Poupouzas
National & Kapodistrian University of Athens
Vassilis Issaris
National & Kapodistrian University of Athens
Konstantina Dede
National & Kapodistrian University of Athens
Helen Christopoulou
General Hospital Attiki KAT
Kyriakos Karkoulias
General Hospital Attiki KAT
Sylvia Raftopoulou
General Hospital of Thoracic Diseases "Sotiria"
Sophia Pouriki
General Hospital of Thoracic Diseases "Sotiria"
Chrysi Keskinidou

National & Kapodistrian University of Athens
Nikolaos S. Lotsios
National & Kapodistrian University of Athens
Spyretta Golemati
National & Kapodistrian University of Athens
Anastasia Kotanidou
National & Kapodistrian University of Athens
Stavros Papapanagiotou
Internet Business Hellas
Ioanna Dimopoulou
National & Kapodistrian University of Athens

Research Article

Keywords: machine learning, sepsis prediction, intensive care unit, adults, prospective study, multicenter study, biomarkers

Posted Date: May 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9703299/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Sepsis is a severe, life-threatening condition arising from a dysregulated host response to infection, where timely recognition is essential for improving patient outcomes. This study aimed to develop and evaluate machine learning (ML) models for early sepsis prediction using non-temporal clinical data collected within the first 24 hours of intensive care unit (ICU) admission, while assessing the contribution of inflammatory and endothelial biomarkers.

Methods

A combined cohort of 1,209 ICU patients from retrospective and prospective multicenter datasets was analyzed. We performed statistical analyses of inflammatory and endothelial biomarkers using chi-square tests of independence. Tree-based models (Random Forest, XGBoost, LightGBM, AdaBoost) and Logistic Regression were evaluated with performance assessed via AUC (Area Under the ROC Curve) and AUPRC (Area Under the Precision-Recall Curve).

Results

LightGBM achieved the best performance, with an AUC of 87.9% and 64.2% AUPRC on our prospective data. Feature ablation demonstrated that inclusion of interleukin (IL)-6, Procalcitonin (PCT), C-Reactive Protein (CRP), and soluble Endothelial Protein C Receptor (sEPCR) significantly improved model performance (AUC: $p = 4.0753 \times 10^{-4}$; AUPRC: $p = 1.6117 \times 10^{-10}$).

Conclusions

These findings demonstrate that accurate early sepsis prediction is feasible using non-temporal data and highlight the critical added value of biologically relevant biomarkers in ML-based clinical decision support.

1. Introduction

Sepsis is a life-threatening syndrome characterized by a dysregulated host response to infection, leading to organ dysfunction and high mortality. According to the current consensus definition introduced by the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [1], sepsis is defined as infection accompanied by acute organ dysfunction, operationalized as an increase in the Sequential Organ Failure Assessment (SOFA) score [2]. Despite advances in critical care, sepsis remains a major global health burden, associated with substantial morbidity, prolonged ICU stays, and significant healthcare costs. Early recognition and timely intervention are critical determinants of survival. Traditional clinical risk stratification tools, such as the Acute Physiology and Chronic Health Evaluation II (APACHE II) [3] and SOFA, are widely used to assess disease severity and estimate mortality risk in critically ill patients. However, these scoring systems were not designed for early sepsis prediction and may not fully capture the complex nonlinear interactions among physiological variables and laboratory markers.

Machine Learning (ML) algorithms are well suited to modeling nonlinear relationships, higher-order feature interactions, and complex temporal dependencies, thus offering the potential for earlier and more accurate prediction in critical illness, including sepsis. However, most ML approaches rely heavily on high-resolution temporal data, which can be difficult to acquire and often necessitate continuous or invasive monitoring protocols. Moreover, many ML

models are predominantly driven by routinely collected vital signs. This reliance is partly attributable to the temporal structure of these variables and to the widespread use of large public datasets such as the Medical Information Mart for Intensive Care (MIMIC)-IV [4], which frequently lack comprehensive representation of established sepsis biomarkers such as procalcitonin (PCT) and C-Reactive Protein (CRP). Evidence from recent systematic reviews supports this observation. Bomrah et al. [5], in a review of 29 studies, identified respiratory rate, heart rate, temperature, diastolic blood pressure, and systolic blood pressure as the most reported predictive features for sepsis development. Similarly, Fleuren et al. [6], in their review of 24 articles, highlighted heart rate, respiratory rate, temperature, systolic blood pressure, and oxygen saturation as the predominant predictors. Islam et al. [7], in a systematic review encompassing 42 studies, reported heart rate, temperature, white blood cell count, systolic blood pressure, and age as the top predictive variables. Collectively, these findings underscore the dominant role of vital signs and the underrepresentation of sepsis biomarkers in contemporary ML-based sepsis prediction models.

Inflammatory biomarkers, including interleukin (IL)-6, PCT, and CRP, have been incorporated into statistical modeling frameworks as predictors of mortality in patients with sepsis [8, 9, 10]. Beyond their prognostic value for mortality, combinations of IL-6, PCT, and CRP have also been utilized to support risk stratification and prediction of sepsis development in adult populations [11]. Soluble Endothelial Protein C Receptor (sEPCR) has also shown potential in sepsis prediction and prognosis in adult populations [12, 13]. Endothelial dysfunction is strongly implicated in the pathophysiology of sepsis [14]. The shedding of sEPCR into the plasma is a direct indicator of endothelial dysfunction and the breakdown of the protein C anticoagulant pathway [15, 17]. While traditional markers such as IL-6 and PCT are highly sensitive to early immune activation, they often lack the specificity needed to distinguish between non-infectious disease states and life-threatening sepsis [16]. Integrating IL-6, CRP, and PCT with sEPCR provides a multidimensional approach to sepsis monitoring by bridging systemic inflammation with functional endothelial assessment [15, 16]. However, the integration of these biomarkers has not been previously studied in ML approaches.

A systematic review of 37 studies revealed that PCT or CRP were reported only in five studies [18]. Liu et al. [19], reported PCT as a top predictive feature using a Random Forests (RF) model, while Yuan et al. [20] using an Extreme Gradient Boosting (XGBoost) model, found CRP as a top predictive feature. However, all studies used temporal data that inherently favor vital signs. To the best of our knowledge the only studies that have not used temporal data are by Yang et al. [21] and Albu et al. [22], who used an XGBoost algorithm or an RF algorithm, respectively. These findings highlight a methodological gap in the literature, concerning early sepsis prediction using the first 24-hour non-temporal data and the systematic evaluation of biomarker contribution. To address this gap, the present study is guided by the following research questions:

- **RQ1**: *What level of predictive performance can be achieved by ML models when trained on clinical data obtained within the first 24 hours of ICU admission for early sepsis prediction?*
- **RQ2**: *What is the relative contribution of inflammatory (IL-6, PCT, CRP) and endothelial (sEPCR) biomarkers to the prediction of sepsis development, and what is their incremental predictive value as determined through systematic feature ablation?*

2. Methods

2.1. Study design

Retrospective data of 601 adult patients admitted to the ICU of “Evangelismos” General Hospital were used, of whom 169 developed sepsis during their ICU stay. Subsequently, prospective data were collected between 17/1/2025 to 19/1/2026 from the ICUs of “Evangelismos” General Hospital, the General Hospital of Thoracic Diseases “Sotiria”, and the General Hospital of Attica “KAT”. The Ethics Committees of the three Hospitals approved the study (study approval

numbers 14–17/01/2025; 27 – 5/02/2025 & 8260-26/03/2025, respectively). Informed consent was obtained from all patients' next-of-kin before inclusion, and all procedures carried out complied with the Helsinki Declaration. The prospective data collected from the 3 ICUs yielded a combined cohort of 608 adult ICU patients, of whom 73 developed sepsis during ICU stay. All patients were followed for 28-day outcomes following ICU admission. Demographics of both datasets are summarized in Table 1. Harmonized data collection protocols were implemented across all participating centers to ensure consistent variable definitions and comparability of datasets.

Table 1
Comparison of demographic characteristics and specific inflammatory and endothelial biomarkers measured on ICU admission, between retrospective and prospective cohorts.

Characteristic	Retrospective	Prospective
ICU admissions	601	608
Sepsis Prevalence (%)	169 (28.12%)	73 (12.01%)
Median age (years)	54	63
Male (%)	65.50%	60.20%
Median day of sepsis development (median, IQR)	6 (5–9)	7 (3–12)
IL-6 (pg/ml), (median, IQR)	96.82 (27.26-233.22)	42.83 (10.24-112.32)
sEPCR (ng/ml) (median, IQR)	124.75 (76.98-191.07)	76.04 (52.50-111.40

The Sepsis-3 definition was applied in this study [1]. Patients who developed sepsis within the first 48 hours after ICU admission were excluded from the study to rule out subjects who might have had subclinical sepsis upon ICU entry.

Both datasets were aligned to comprise 43 shared variables, including demographic and clinical characteristics (e.g., age, sex, admission diagnosis), vital signs (e.g., temperature, heart rate, mean arterial pressure, respiratory rate, and oxygenation/ventilatory parameters), routine laboratory biomarkers (e.g., PCT, CRP, lactate, blood counts, coagulation indices, renal and liver function tests, and blood gas parameters), specific inflammatory and endothelial biomarkers (IL-6 and sEPCR), and treatment-related variables (mechanical ventilation on admission and vasopressor/inotrope use).

2.2. Measurement of IL-6 and sEPCR

Interleukin (IL)-6 and sEPCR were measured in the serum samples of all retrospective and prospective patients via immunoassays purchased from Invitrogen (Thermo Fisher Scientific Inc. Waltham, MA, USA; detection limits: IL-6, < 2 pg/ml and sEPCR, 0.035 ng/ml), according to the manufacturer's instructions.

2.3. Model development and validation

Tree-based models (RF, XGBoost, Light Gradient Boosting Machine (Light- GBM), Adaptive Boosting (AdaBoost)) and Logistic Regression were trained and evaluated on both retrospective and prospective cohorts. Tree- based methods were selected given their strong performance in related studies, whereas logistic regression was included as a conventional baseline for comparison. Given the limited sample sizes, performance was estimated across 50 repeated train–test splits during model training, and across 10 splits during external validation. Model performance was evaluated using Area Under the ROC Curve (AUC) and Area Under the Precision Recall Curve (AUPRC).

2.4. Ablation study design

To quantify the incremental contribution of selected biomarker groups, we conducted a structured feature ablation study. Predictors were organized into clinically meaningful categories, and multiple model variants were constructed

accordingly: *(i)* the complete model including all candidate predictors, *(ii)* excluding inflammatory and endothelial markers (IL-6, PCT, CRP, sEPCR), *(iii)* excluding vital signs (Heart Rate, Temperature, Respiratory Rate, Mean Arterial Pressure). The complete and ablated models were trained using identical preprocessing, data splits, and hyperparameter tuning procedures.

2.5. Statistical analysis

Paired hypothesis testing was performed to compare model performance across repeated data splits, using performance metrics computed on the same splits for each model. Specifically, paired *t*-tests were used to evaluate whether the mean performance difference between models and scoring systems deviated from zero. Differences were considered statistically significant at $p < 0.05$. To explore the robustness of the association between the set of inflammatory and endothelial biomarkers measured and sepsis status, a threshold-scanning approach was performed. For each biomarker, patients were dichotomized according to whether their biomarker level exceeded a series of candidate thresholds. At each threshold, a chi-square test of independence was conducted to evaluate the association between exceeding the threshold and sepsis status. To account for multiple hypothesis testing across thresholds, p-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure.

2.6. Model interpretation

SHAP (SHapley Additive exPlanations) [23] was used to generate feature importance, Beeswarm, and dependence plots to elucidate model predictions.

2.7. Software

Our ML models were developed and trained using Python 3.12. Experiments were conducted on an Intel Core i7-14700k CPU and RTX 5060 GPU.

3. Results

3.1. Comparison of ML algorithms

We trained and evaluated four tree-based models (RF, XGBoost, LightGBM, and AdaBoost) alongside a linear baseline model (Logistic Regression). The comparative performance of these models is presented in Table 2. Of importance, the tree-based approaches consistently outperformed Logistic Regression across all evaluation metrics, with LightGBM achieving the highest performance in both AUC and AUPRC.

Table 2: Performance metrics (AUC and AUPRC) for all models, evaluated on both retrospective and prospective cohorts. LightGBM outperformed other models in both evaluation metrics across cohorts.

	Retrospective		Prospective	
Model	**AUC**	**AUPRC**	**AUC**	**AUPRC**
LR	0.679(0.575,0.770)	0.474(0.366,0.605)	0.749(0.650,0.883)	0.449(0.282,0.695)
RF	0.848(0.783,0.920)	0.655(0.515,0.806)	0.868(0.760,0.947)	0.589(0.343,0.849)
XGBoost	0.842(0.778,0.909)	0.637(0.499,0.786)	0.870(0.766,0.957)	0.603(0.346,0.871)
LightGBM	0.849(0.788,0.912)	0.666(0.527,0.787)	0.879(0.781,0.949)	0.642(0.349,0.857)
AdaBoost	0.828(0.756,0.898)	0.634(0.493,0.772)	0.835(0.706,0.932)	0.525(0.331,0.778)

During external validation from the retrospective (source) to the prospective (target) cohort, LightGBM achieved an AUC of 0.758 (95% CI: (0.719,0.799)) and an AUPRC of 0.345 (95% CI: (0.263,0.445)).

3.2. Ablation study on IL-6, PCT, CRP, and sEPCR

Table 3 presents a comprehensive summary of the findings from the feature ablation analysis. Figures 1 and 2 present the performance of LightGBM evaluated with and without our set of inflammatory and endothelial biomarkers and compared to the APACHE II and SOFA scoring systems. The removal of these biomarkers resulted in a decrease of approximately 2% in AUC and 7% in AUPRC. Moreover, the complete LightGBM model demonstrated statistically significant performance improvements compared to the ablated model, as confirmed by paired testing (AUC: $p = 4.0753 \times 10^{-4}$; AUPRC: $p = 1.6117 \times 10^{-10}$).

Table 3: Performance metrics (AUC and AUPRC) for LightGBM in the prospective cohort during feature ablation analysis.

	Prospective	
Model	**AUC**	**AUPRC**
LightGBM	0.879(0.781,0.949)	0.642(0.349,0.857)
without inflammatory and endothelial biomarkers	0.864(0.765,0.943)	0.573(0.345,0.796)
without vital signs	0.876(0.765,0.947)	0.640(0.353,0.839)

3.3. Statistical analysis of inflammatory and endothelial biomarkers

IL-6 demonstrated a strong and consistent association with sepsis status across a wide range of thresholds (Fig. 3a). Of the 30 evaluated thresholds, 24 (80%) remained significant after FDR correction ($q < 0.05$). The strongest association was observed at a threshold of 320 ($p = 3.777 \times 10^{-18}$, $q = 1.133 \times 10^{-16}$), representing the minimum q-value across tested cutoffs. Significant associations persisted across most thresholds, indicating robust separation between septic and non-septic patients over a broad concentration range. PCT similarly showed strong threshold-dependent associations with sepsis (Fig. 3b). Of 16 tested thresholds, 15 (94%) remained significant after FDR correction. The strongest association was observed at a threshold of 1.146 ($p = 3.069 \times 10^{-13}$, $q = 4.912 \times 10^{-12}$). Significant thresholds spanned approximately 0.6 to 8.5, suggesting a stable relationship between elevated PCT levels and sepsis status. In contrast, neither sEPCR nor CRP demonstrated statistically significant associations after FDR correction (Figs. 3d, 3c). Although some raw p-values were suggestive, no tested thresholds remained significant after adjustment (0/39 for sEPCR and 0/7 for CRP). These findings indicate weaker or less consistent threshold-dependent discrimination for these biomarkers.

3.4. Explainability of predictors

Figure 4 presents the top 20 features ranked by their mean Gini importance values for LightGBM, computed over 50 independent shuffles to ensure robustness. Among the examined biomarkers, IL-6 emerged as the most influential predictor, demonstrating the highest mean importance across all variables. CRP and PCT also ranked prominently within the top predictors, confirming their strong contribution to model discrimination. Notably, sEPCR retained meaningful predictive value, although with lower relative importance compared to IL-6, CRP, and PCT.

The SHAP Beeswarm plot indicated that inotropic/vasopressor support and IL-6 are the most influential predictors in the prospective cohort (Fig. 5). High IL-6 levels are consistently associated with positive SHAP values, indicating an increased predicted probability of sepsis. Similarly, elevated PCT and CRP contribute positively to model output,

reinforcing their role as key inflammatory biomarkers. Notably, sEPCR also demonstrates a moderate contribution, with higher values tending to shift predictions toward increased sepsis risk. To further investigate nonlinear relationships between biomarkers and model predictions, SHAP dependence plots were generated for IL-6, PCT, CRP, and sEPCR (Fig. 6). IL-6 and PCT demonstrated a pronounced nonlinear threshold-like pattern. At low concentrations (< 100 pg/ml), SHAP values for IL-6 were near zero or slightly negative, however, a marked increase in SHAP values was observed around 150–200 pg/ml (Fig. 6a), after which IL-6 exerted a strong positive effect on predicted sepsis risk. This behavior suggests the presence of a biologically meaningful inflection point. Similarly, for PCT, SHAP values increased sharply at approximately 1.5 ng/ml (Fig. 6b). CRP and sEPCR exhibited non-monotonic relationships with the model output. Specifically, lower CRP values were still associated with increased SHAP values (Fig. 6c), while lower sEPCR levels did not consistently correspond to lower SHAP values (Fig. 6d).

4. Discussion

In this multicenter study, we developed and evaluated various ML models for the prediction of sepsis development in adult critically ill patients. Among these, LightGBM achieved the best overall performance (AUC = 88%, AUPRC = 64%). The inclusion of inflammatory and endothelial biomarkers (IL-6, PCT, CRP, and sEPCR) significantly enhanced model performance, as evidenced by the observed decline in both AUC (2%) and AUPRC (7%) following their removal, with these differences confirmed by paired statistical testing. In comparison, when performing model ablation by removing vital signs (Heart Rate, Temperature, Respiratory Rate, Mean Arterial Pressure) the model exhibited a decrease of only 0.3% AUC and 0.2% AUPRC. Both the LGBM complete and ablated models outperformed the APACHE II and SOFA scoring systems. Our approach generalized well, considering the limited sample size, achieving an AUC of 76% during external validation.

The threshold-based statistical analysis further highlighted the differential discriminatory capacity of the evaluated biomarkers. IL-6 demonstrated a robust and consistent association with sepsis across a wide range of thresholds, with the majority remaining significant after FDR correction, indicating strong and stable separation between septic and non-septic patients. Similarly, PCT exhibited a highly consistent threshold-dependent relationship, with nearly all evaluated cutoffs retaining statistical significance after multiple testing correction, supporting its role as a reliable biomarker across clinically relevant concentration ranges. In contrast, neither sEPCR nor CRP showed statistically significant associations following FDR adjustment, despite suggestive unadjusted p-values. This lack of significance after correction suggests that their discriminatory ability is weaker, less stable, or potentially context dependent. Our findings suggest that IL-6 and PCT provide more robust and reproducible signal for sepsis discrimination compared to CRP and sEPCR.

The SHAP analysis provided important insights into the nonlinear effects of key predictors within the model. Inotropic/vasopressor support and IL-6 emerged as the most influential features, underscoring the combined importance of hemodynamic instability and systemic inflammation in sepsis prediction. Consistent with their known biological roles, higher levels of IL-6, PCT, and CRP were generally associated with increased predicted risk. Notably, IL-6 and PCT exhibited clear nonlinear, threshold-like behavior, suggesting the presence of biologically meaningful inflection points. In contrast, CRP and sEPCR demonstrated non-monotonic relationships, indicating more complex and less predictable effects on model output. Specifically, lower values of these biomarkers did not consistently correspond to reduced predicted risk, highlighting potential context-dependent or heterogeneous roles in sepsis pathophysiology.

Consistent with prior systematic reviews [5, 6, 7, 18], ML models have relied predominantly on vital signs and routinely collected clinical variables, often derived from large temporal datasets such as MIMIC-IV. In contrast, our approach demonstrates that competitive predictive performance can be achieved using non-temporal data from the first 24 hours of ICU admission, while explicitly incorporating biologically relevant inflammatory and endothelial biomarkers.

Our results highlight the substantial contribution of IL-6 and PCT, that have been systematically studied using statistical methods but not ML methods [8, 9, 10]. Furthermore, the observed performance degradation following biomarker ablation provides quantitative evidence of their incremental predictive value, addressing a gap in related work where such contributions have often been assumed but not rigorously tested. To the best of our knowledge, only the studies by Yang et al. [21] and Albu et al. [22] have adopted similar non-temporal frameworks, utilizing XGBoost and RF models, respectively. While both studies demonstrated the feasibility of sepsis prediction without high-resolution time-series data, they did not systematically evaluate the contribution of inflammatory and endothelial biomarkers. In contrast, our study provides quantitative evidence of the incremental value of biomarkers such as IL-6, PCT, CRP, and sEPCR through a structured ablation analysis. Furthermore, our use of SHAP-based interpretability enables a more detailed characterization of non-linear and threshold-dependent biomarker effects, which were not explored in these prior works. Our findings differ from those reported by Yang et al. [21], who identified CRP as a statistically significant predictor using chi-squared testing, whereas in our analysis, statistical significance after multiple testing correction was observed for PCT but not for CRP. This discrepancy may be attributable to differences in cohort characteristics.

This study has several important strengths. First, the inclusion of a prospective, multicenter cohort enhances the clinical relevance of the findings, as it reflects real-world variability in patient populations and clinical practice. Second, the integration of biologically meaningful inflammatory and endothelial biomarkers, including IL-6, PCT, and sEPCR, represents a key contribution, as these markers remain underrepresented in ML–based sepsis prediction studies, which have traditionally focused on routinely collected vital signs. Third, the study design allowed for a systematic evaluation of biomarker contribution through feature ablation, providing quantitative evidence of their incremental predictive value. Finally, the use of SHAP-based interpretability enabled detailed characterization of nonlinear and non-monotonic relationships, offering insights into the underlying biological patterns captured by the model and improving the transparency of predictions.

Several limitations of this study should be acknowledged. First, although both retrospective and prospective cohorts were included, the overall sample size remains modest for ML applications, where substantially larger datasets are typically required to ensure model stability and generalizability. Second, the use of non-temporal data limits the ability of the model to capture dynamic changes in biomarkers and clinical variables over time, which may carry important prognostic information, particularly in a rapidly evolving condition such as sepsis. Finally, while external validation was performed, the partial overlap of participating centers between the retrospective and prospective cohorts may have introduced shared institutional patterns, potentially leading to an overestimation of model generalizability. Future studies incorporating larger, fully independent, and temporal datasets are warranted to address these limitations.

5. Conclusions

In conclusion, this study demonstrates that early prediction of sepsis can be achieved using ML models trained on non-temporal clinical data obtained within the first 24 hours of ICU admission that is competitive with ML models trained on temporal datasets. Among the evaluated approaches, LightGBM achieved the best overall performance and maintained consistent discrimination during external validation. Importantly, the integration of inflammatory and endothelial biomarkers, particularly IL-6 and PCT, provided significant incremental predictive value, as confirmed through systematic ablation and statistical analysis. SHAP-based interpretation further revealed clinically meaningful nonlinear and threshold-dependent relationships, highlighting the complex role of these biomarkers in sepsis pathophysiology. These findings support the development of biologically informed and interpretable ML models as a promising approach for early sepsis prediction. Future work should focus on validating these findings in larger, fully independent cohorts and incorporating temporal data to further enhance predictive performance and clinical applicability.

Declarations

Data and code availability

The code used to reproduce our experiments is available in the Github repository: https://github.com/giannis3p/sepsis-biomarker-ml. The datasets supporting the conclusions of this article are available from the authors upon reasonable request.

Author Contributions

IP: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. AGV: Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. CSV: Writing – review & editing, Visualization, Validation, Methodology, Investigation, Data curation. MT: Writing – review & editing, Visualization, Validation, Methodology, Investigation. ZM: Writing – review & editing, Visualization, Validation, Methodology, Investigation. AKa: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. VG.: Writing – original draft, Methodology, Investigation, Data curation. OK: Writing – original draft Methodology, Investigation, Data curation. IC: Writing – review & editing, Visualization, Validation, Methodology, Investigation. GP: Writing – original draft, Methodology, Investigation, Data curation. VI: Writing – original draft, Methodology, Investigation, Data curation. KD: Writing – original draft, Methodology, Investigation, Data curation. HC: Writing – original draft, Methodology, Investigation, Data curation. KK: Writing – original draft, Methodology, Investigation, Data curation. SR: Writing – original draft, Methodology, Investigation, Data curation. SP: Writing – original draft, Methodology, Investigation, Data curation. CK: Writing – original draft, Methodology, Investigation, Data curation. NSL: Writing – original draft, Methodology, Investigation, Data curation. SG.: Writing – review & editing, Visualization, Validation, Methodology, Investigation. AKo: Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. SP: Formal analysis, Investigation, Methodology, Software, Validation, Writing – review & editing. ID: Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Funding

The work was implemented under the Clusters of Research Excellence (CREs), funded by the European Union, Recovery and Resilience Facility (RRF), Greece 2.0, Next Generation EU. Grant number: ΥΠ3ΤΑ-0559412.

Competing Interests

The authors declare that they have no competing interests.

Ethical Approval and Consent to participate

The Ethics Committees of the three Hospitals approved the study (study approval numbers 14-17/01/2025; 27-5/02/2025 & 8260-26/03/2025, respectively). Informed consent was obtained from all patients' next-of-kin before inclusion, and all procedures carried out complied with the Helsinki Declaration.

Consent for publication

All authors have approved the submitted version and have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in

which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Acknowledgements

Not applicable.

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche J-D, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent J-L, Angus DC. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA. 2016;315(8):801–10.
2. Vincent J-L, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22(7):707–10.
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.
4. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, Pollard TJ, Hao S, Moody B, Gow B, Lehman L-WH, Celi LA, Mark RG. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1.
5. Bomrah S, Uddin M, Upadhyay U, Komorowski M, Priya J, Dhar E, Hsu S-C, Syed-Abdul S. A scoping review of machine learning for sepsis prediction- feature engineering strategies and model performance: a step towards explainability. Crit Care. 2024;28(1):180. 10.1186/s13054-024-04948-6.
6. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, Swart EL, Girbes ARJ, Thoral P, Ercole A, Hoogendoorn M, Elbers PWG. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med. 2020;46(3):383–400. 10.1007/s00134-019-05872-y.
7. Islam KR, Prithula J, Kumar J, Tan TL, Reaz MBI, Sumon MSI, Chowdhury MEH. Machine learning-based early prediction of sepsis using electronic health records: A systematic review. J Clin Med. 2023;12(17):5658.
8. Yu B, Chen M, Zhang Y, Cao Y, Yang J, Wei B, Wang J. Diagnostic and prognostic value of interleukin-6 in emergency department sepsis patients, Infect. Drug Resist. 2022;15:5557–66.
9. Xie Y, Zhuang D, Chen H, Zou S, Chen W, Chen Y. 28-day sepsis mortality prediction model from combined serial interleukin-6, lactate, and procalcitonin measurements: a retrospective cohort study. Eur J Clin Microbiol Infect Dis. 2023;42(1):77–85. 10.1007/s10096-022-04517-1.
10. Tschaikowsky K, Hedwig-Geissing M, Braun GG, Radespiel-Troeger M. Predictive value of procalcitonin, interleukin-6, and c-reactive protein for survival in postoperative patients with severe sepsis. J Crit Care. 2011;26(1):54–64. https://doi.org/10.1016/j.jcrc.2010.04.011.
11. Carcò D, Castorina P, Guardo P, Iachelli V, Pace T, Scirè P, Stanco R, Markovic U, Milone GA, Galbo F, Moschetti G, Martorana E. Combination of interleukin-6, c-reactive protein and procalcitonin values as predictive index of sepsis in course of fever episode in adult haematological patients: Observational and statistical study. J Clin Med. 2022;11(22). 10.3390/jcm11226800.
12. Vassiliou AG, Kotanidou A, Mastora Z, Maniatis NA, Albani P, Jahaj E, Koutsoukou A, Armaganidis A, Orfanos SE. Elevated soluble endothelial protein C receptor levels at ICU admission are associated with sepsis development. Minerva Anestesiol. 2014;81(2):125–34.

13. Guitton C, Gérard N, Sébille V, Bretonnière C, Zambon O, Villers D, Charreau B. Early rise in circulating endothelial protein C receptor correlates with poor outcome in severe sepsis. Intensive Care Med. 2011;37(6):950–6. 10.1007/s00134-011-2171-y. . Epub 2011 Mar 11. PMID: 21394629; PMCID: PMC3529933.
14. Tang F, Zhao X-L, Xu L-Y, Zhang J-N, Ao H, Peng C. Endothelial dysfunction: Pathophysiology and therapeutic targets for sepsis-induced multiple organ dysfunction syndrome. Biomed Pharmacother. 2024;178. https://doi.org/10.1016/j.biopha.2024.117180.
15. Kager LM, Schouten M, Wiersinga WJ, de Boer JD, Lattenist LCW, Roelofs JJTH, Meijers JCM, Levi M, Dondorp AM, Esmon CT, van 't Veer C, van der Poll T. Overexpression of the endothelial protein c receptor is detrimental during pneumonia-derived gram-negative sepsis (melioidosis). PLoS Negl Trop Dis. 2013;7(7):1–11. 10.1371/journal.pntd.0002306.
16. Liang Y, Huang X, Jiang Y, Qin Y, Peng D, Huang Y, Li J, Sooranna SR, Pinhu L. Endothelial protein c receptor polymorphisms and risk of sepsis in a chinese population, Journal of International Medical Research 45 (2) (2017) 504–513, pMID: 28415941. 10.1177/0300060516686496
17. O'Hehir ZD, Lynch T, O'Neill S, March L, Xue M. Endothelial protein c receptor and its impact on rheumatic disease. J Clin Med. 2024;13(7). 10.3390/jcm13072030.
18. Papapanagiotou I, Karalis A, Kokkoris S, Vrettou CS, Kampouropoulou O, Giannopoulou V, Golemati S, Kotanidou A, Papapanagiotou S, Dimopoulou I, Vassiliou AG. Machine learning for early detection and prediction of sepsis: explainability. Int J Med Informatics. 2026;214. https://doi.org/10.1016/j.ijmedinf.2026.106420.
19. Xi Liu M, Li X, Liu Y, Luo D, Yang H, Ouyang J, He J, Xia F, Xiao. Clinical validation and optimization of machine learning models for early prediction of sepsis. Front Med Volume. 2025;12–2025. 10.3389/fmed.2025.1521660.
20. Yuan K-C, Tsai L-W, Lee K-H, Cheng Y-W, Hsu S-C, Lo Y-S, Chen R-J. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. Int J Med Informatics. 2020;141:104176. https://doi.org/10.1016/j.ijmedinf.2020.104176.
21. Yang AC, Ma W-M, Chiang D-H, Liao Y-Z, Lai H-Y, Lin S-C, Liu M-C, Wen K-T, Lin T-H, Tsai W-X, Zhu J-D, Chen T-Y, Lee H-F, Liao P-H, Yien H-W, Wang C-Y. Early prediction of sepsis using an xgboost model with single time-point non-invasive vital signs and its correlation with c-reactive protein and procalcitonin: A multi-center study. Intelligence-Based Med. 2025;11:100242. https://doi.org/10.1016/j.ibmed.2025.100242.
22. Albu A, Holke F, Cao H, Hahn B, Lindner HA, Feuerbach L, Schneider-Lindner V. Sepsis prediction on icu admission: Limits and opportunities for machine learning. J Crit Care. 2025;86:154973. https://doi.org/10.1016/j.jcrc.2024.154973.
23. Lundberg S, Lee S-I. A unified approach to interpreting model predictions (2017). arXiv:1705.07874.

Figures

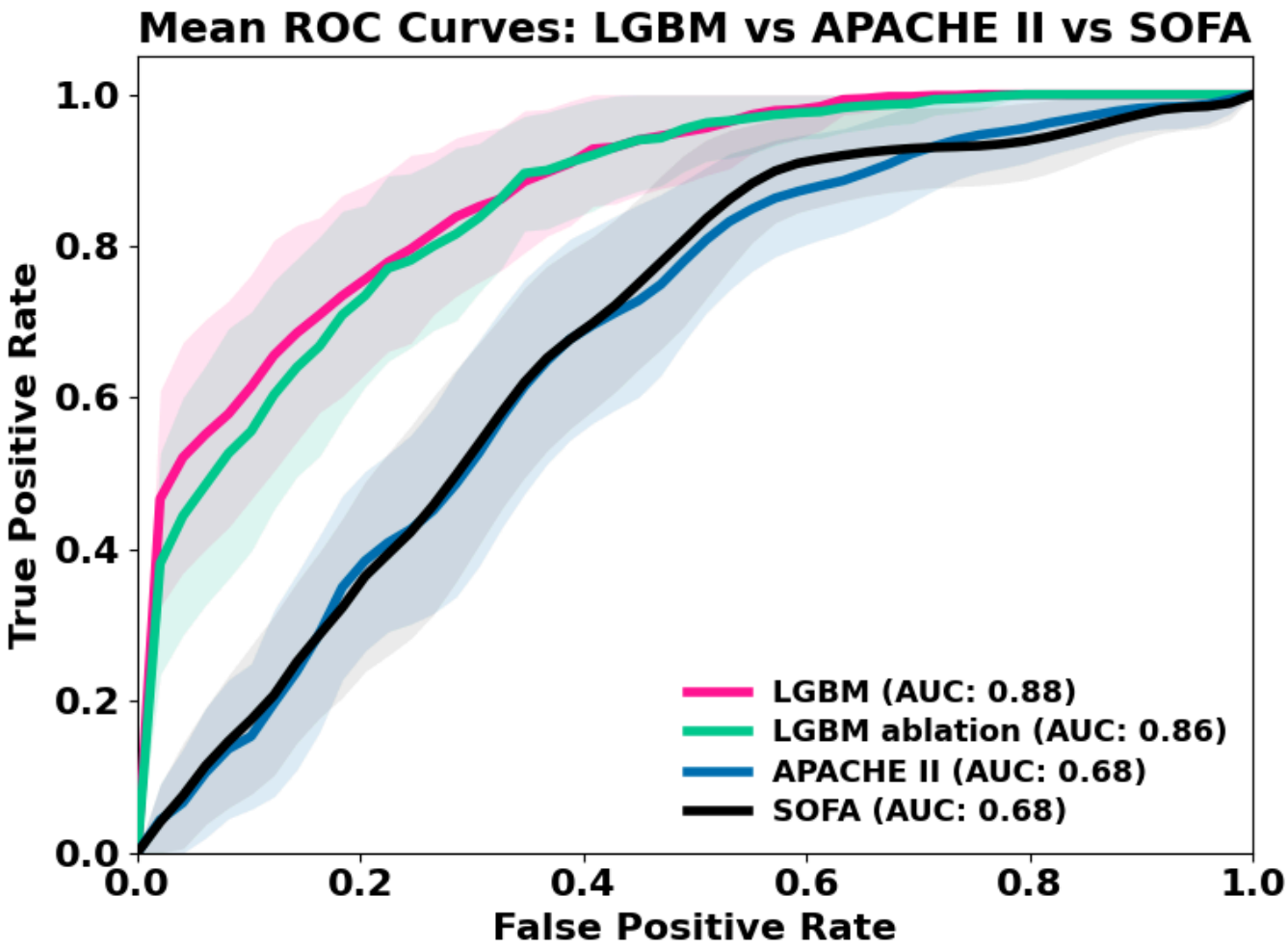


Figure 1

Mean ROC curves for LightGBM, LightGBM ablation, APACHE II, and SOFA. Solid lines represent the mean across 50 splits, and shaded regions show ±1 standard deviation. LGBM (pink line) represents the complete model, and LGBM ablation (green line) represents the LGBM model without inflammatory and endothelial biomarkers.

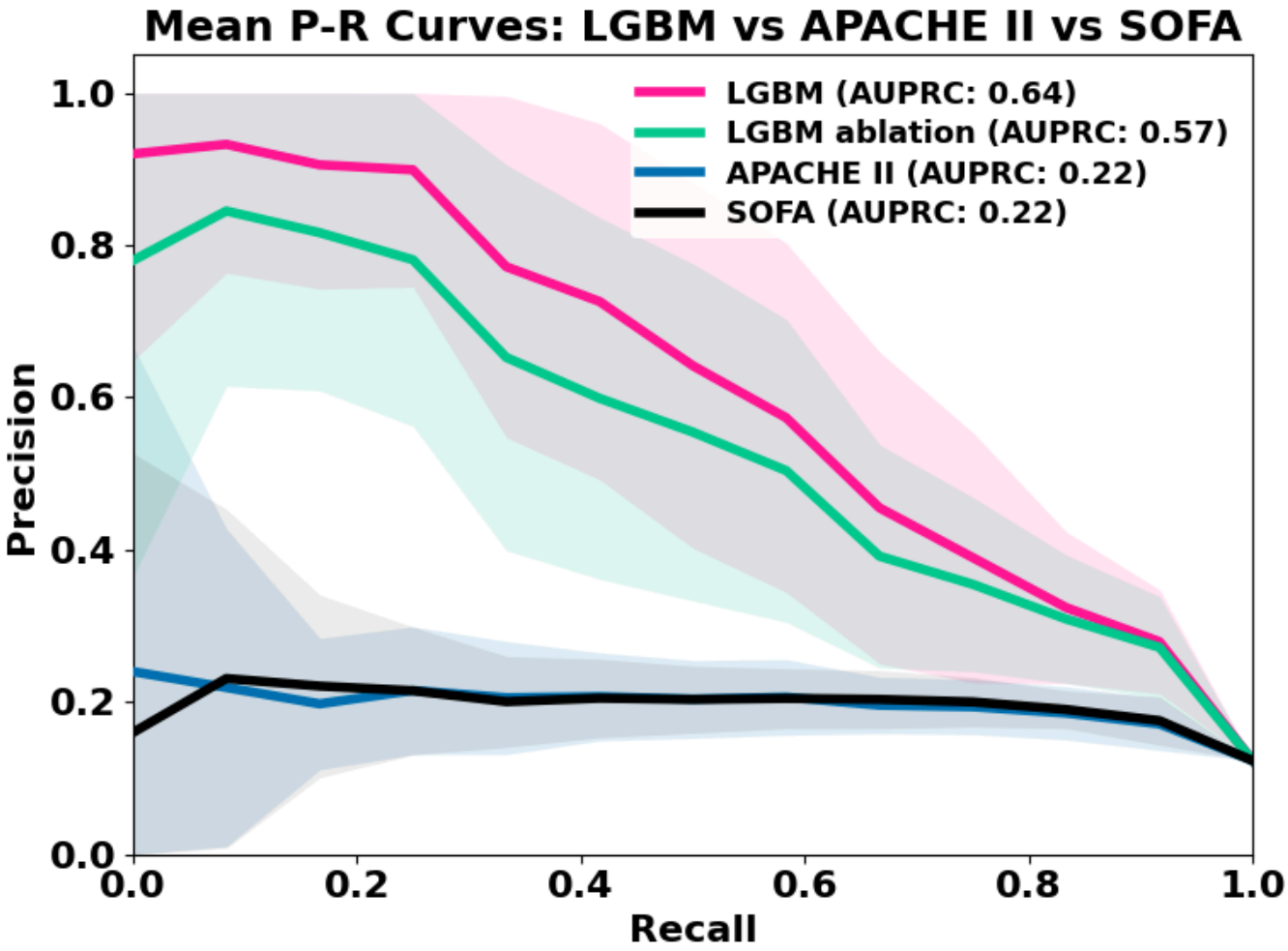


Figure 2

Mean P-R curves for LightGBM, LightGBM ablation, APACHE II, and SOFA. Solid lines represent the mean across 50 splits, and shaded regions show ±1 standard deviation. LGBM (pink line) represents the complete model, and LGBM ablation (green line) represents the LGBM model without inflammatory and endothelial biomarkers.

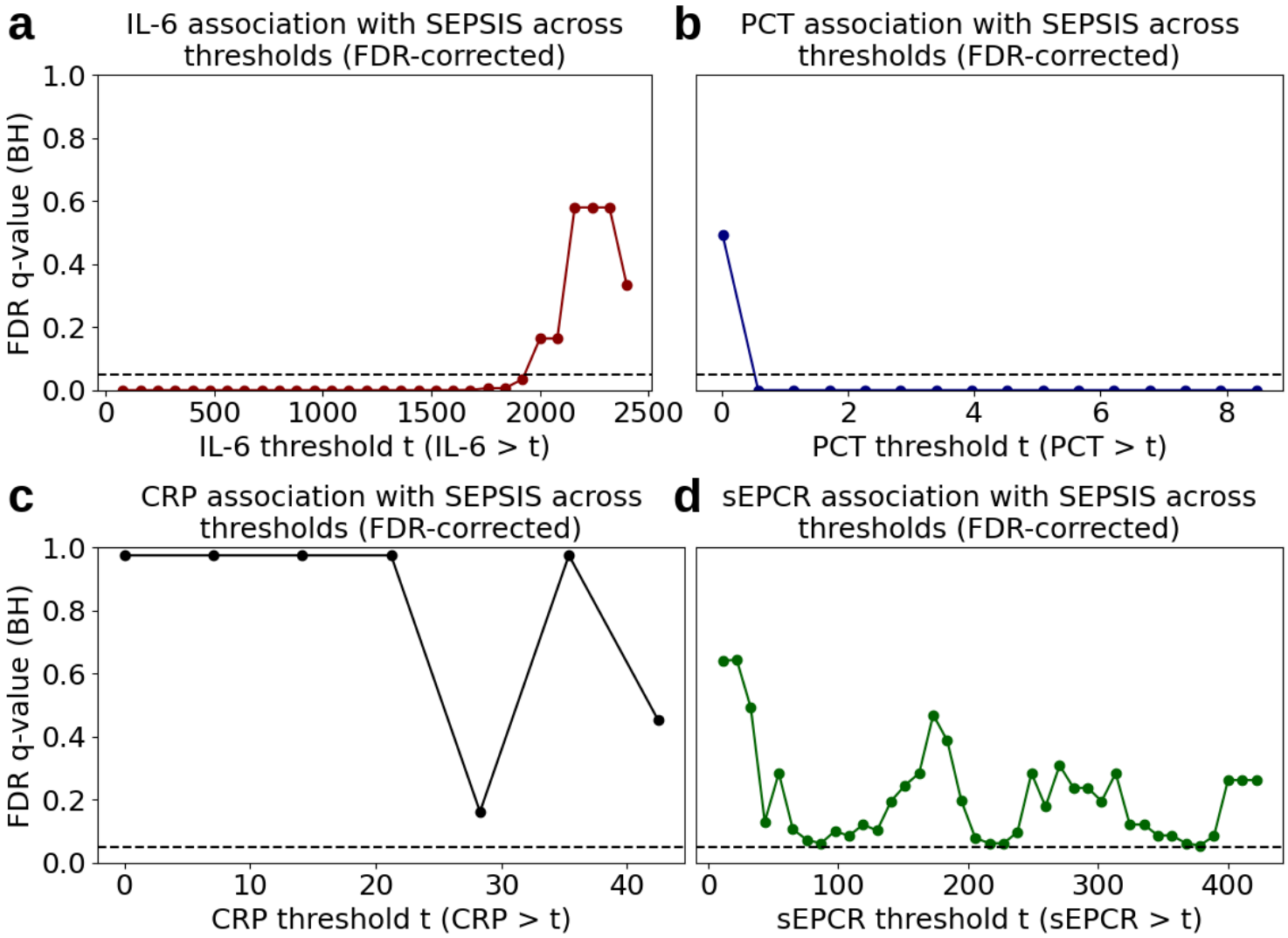


Figure 3

Threshold-based association analysis between inflammatory and endothelial biomarkers and sepsis status. For each biomarker, chi-square tests were performed across a range of thresholds *t*. Raw p-values were adjusted for multiple threshold testing using the Benjamini–Hochberg false discovery rate (FDR) procedure. The dashed horizontal line indicates the significance level (q = 0.05). IL-6 (a) and PCT (b) demonstrated significant associations across multiple thresholds after FDR correction, whereas CRP (c) and sEPCR (d) did not retain statistical significance.

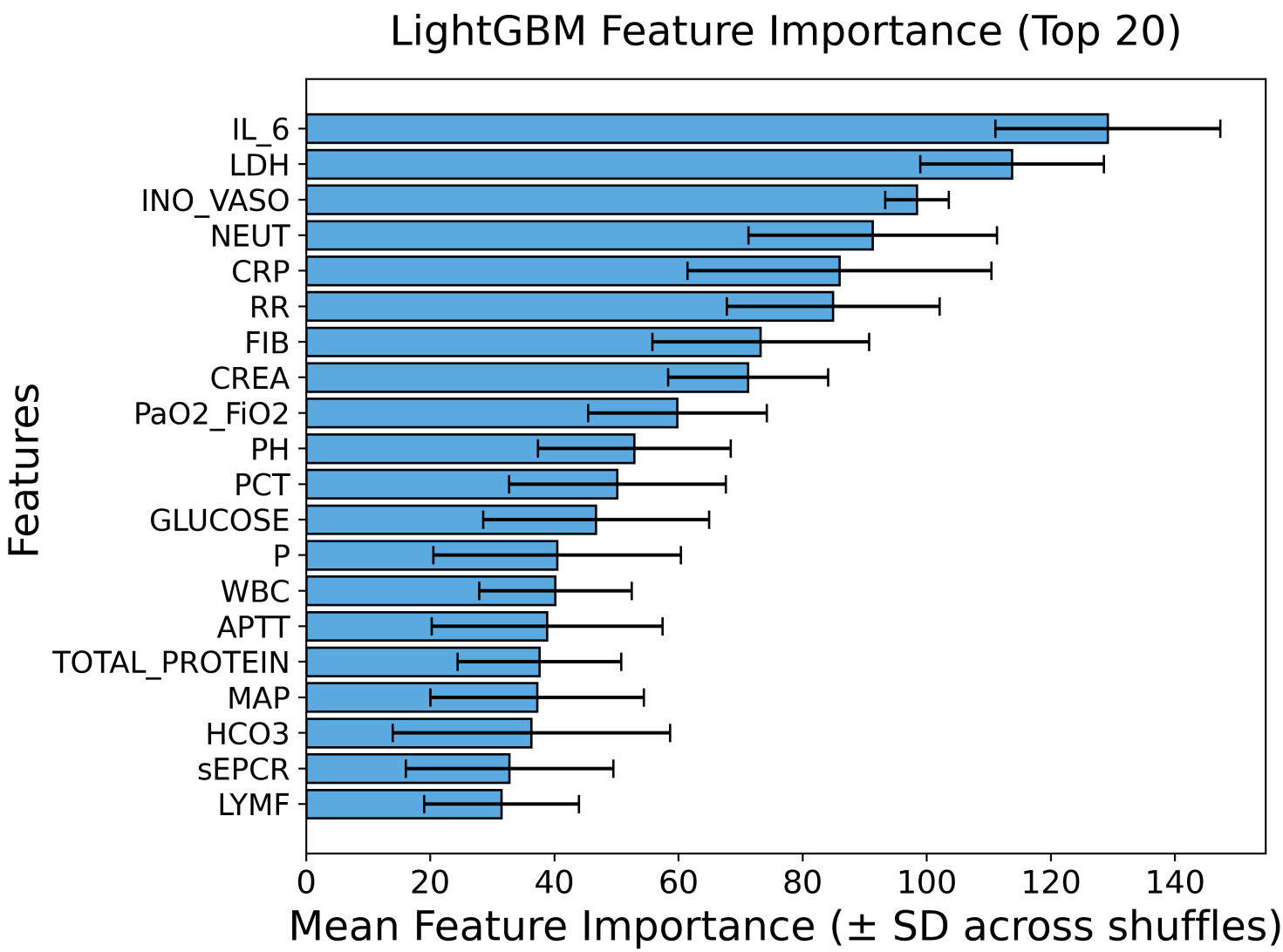


Figure 4

Gini feature importances for LightGBM on the prospective data. *Abbreviations: IL-6 = Interleukin-6; PCT = Procalcitonin; LYMF = Lymphocytes percentage; RR = Respiratory Rate; LDH = Lactate Dehydrogenase; FIB = FIBRINOGEN; APTT = Activated Partial Thromboplastin Time; NEUT = Neutrophils percentage; CREA = Creatinine; PaO2/FiO2 = Arterial Oxygen Partial Pressure to Fraction of Inspired Oxy- gen ratio; sEPCR = soluble Endothelial Protein C Receptor; CRP = C-Reactive Protein; WBC = White Blood Cell count; INO/VASO = Inotropes/Vasopressors; MAP = Mean Arterial Pressure; P = Phosphate; HC03 = Bicarbonate.*

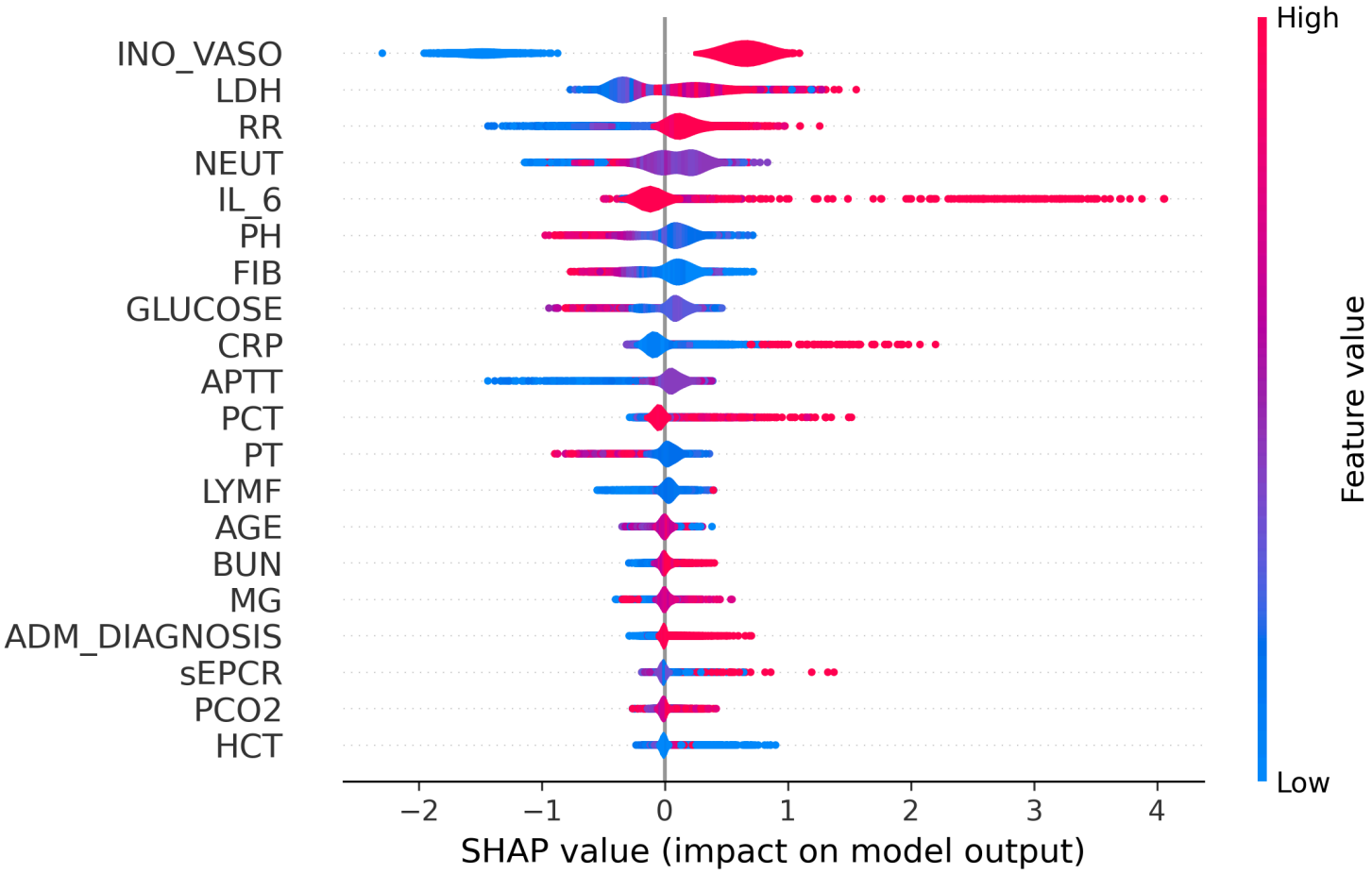


Figure 5

SHAP Beeswarm plot for LightGBM evaluated on the prospective cohort. Features are ranked on the y-axis according to their mean absolute SHAP value, reflecting their overall contribution to model predictions. Each point represents an individual patient sample. Positive SHAP values correspond to an increased predicted probability of sepsis. The color gradient represents the original feature value. *Abbreviations: IL-6 = Interleukin- 6; PCT = Procalcitonin; LYMF = Lymphocytes percentage; RR = Respiratory Rate; LDH = Lactate Dehydrogenase; FIB = FIBRINOGEN; APTT = Activated Partial Thromboplastin Time; PCO2 = Partial Pressure of Carbon Dioxide; BUN = Blood Urea Nitrogen; MG = Magnesium; NEUT = Neutrophils percentage; PT = Prothrombin Time; HCT = Hematocrit; ADM_DIAGNOSIS = Admission Diagnosis; sEPCR = soluble Endothelial Protein C Receptor; CRP = C-Reactive Protein; INO/VASO = Inotropes/Vasopressors.*

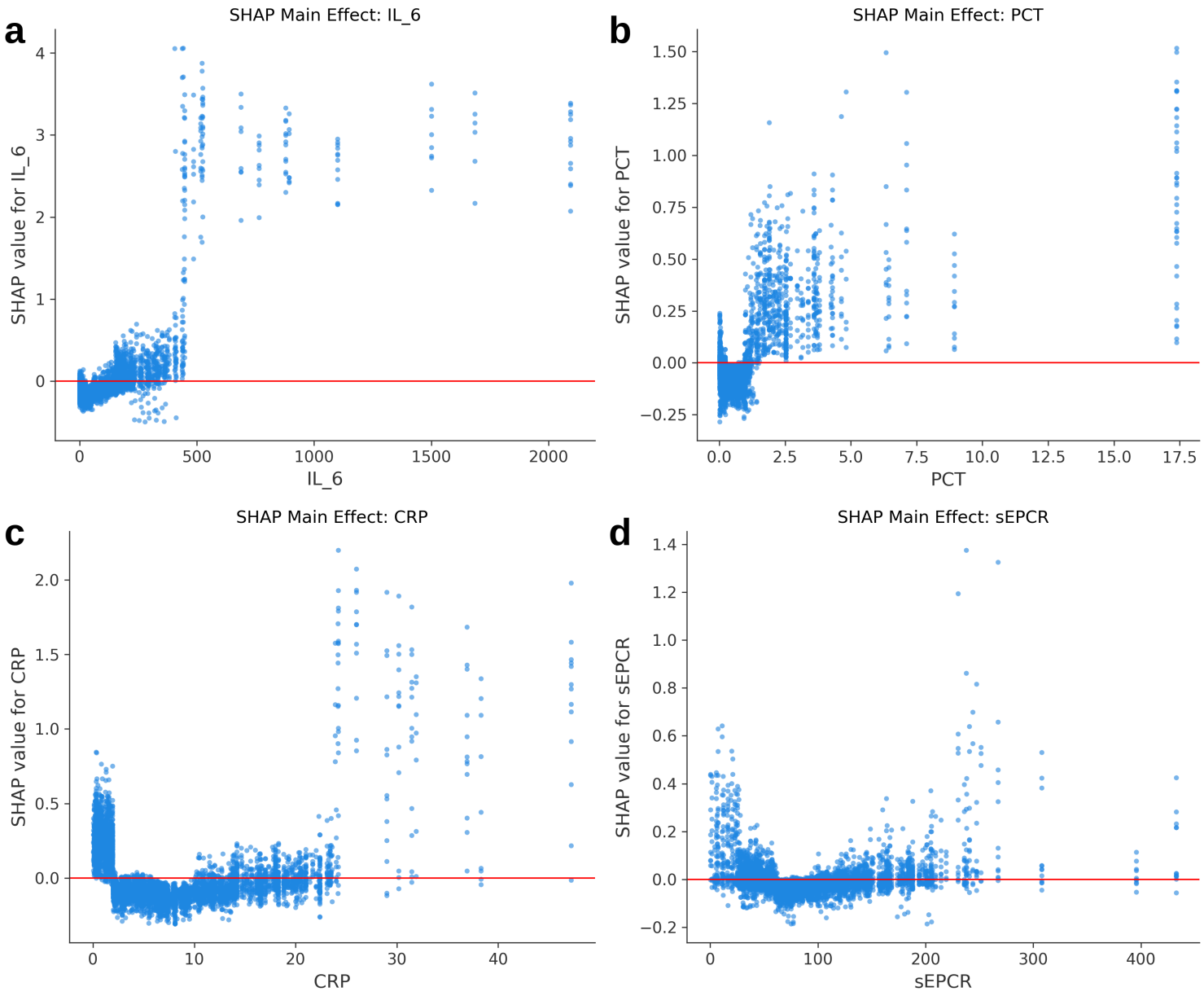


Figure 6

SHAP dependence plots for (a) IL-6, (b) PCT, (c) CRP, and (d) sEPCR derived from LightGBM on the prospective cohort. Each point represents an individual patient. The x-axis denotes the biomarker value, and the y-axis represents the SHAP value (impact on model output). Positive SHAP values increase the predicted probability of sepsis, whereas negative values decrease it. The red horizontal line indicates zero contribution.