JURNAL LOCUS: Penelitian & Pengabdian
Volume 5 No. 11 November 2026, pp 10647-10657
E-ISSN 2829-7334
DOI: 10.58344/locus.v5i11.6453

Identification of Diagnostic Biomarkers and Immune-Related Molecular Signatures in Sepsis Through Integrated Transcriptomic and Machine Learning Analysis

William, Nelson Sudiyono, Ian Pranandi*
Universitas Katolik Indonesia Atma Jaya, Indonesia
Email : william@atmajaya.ac.id, nelson.sudiyono@atmajaya.ac.id, ian.pranandi@atmajaya.ac.id*

Keywords:
Sepsis; Transcriptomics; Diagnostic Biomarkers; Machine Learning; Immune-Related Genes.

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Riviera Publishing

ABSTRACT

Sepsis is a life-threatening syndrome characterized by a dysregulated host response to infection. Its molecular heterogeneity poses a significant challenge to early diagnosis, highlighting the need for biomarkers that reflect sepsis-associated immune alterations. This study aimed to identify immune-related diagnostic biomarkers for sepsis through integrated transcriptomic and machine learning analyses. Two independent whole-blood transcriptomic datasets comprising 190 samples were analyzed: a discovery dataset consisting of 80 patients with sepsis and 40 healthy controls and a validation dataset consisting of 45 patients with sepsis and 25 healthy controls. Differentially expressed genes (DEGs) were identified and intersected with a curated set of immune-related genes. Functional enrichment analysis was performed to characterize the resulting immune-related DEGs (IR-DEGs). Candidate diagnostic biomarkers were selected using least absolute shrinkage and selection operator (LASSO) logistic regression and random forest analysis. A combined five-gene model was developed, and its diagnostic performance was assessed using receiver operating characteristic (ROC) curve analysis. A total of 1,246 DEGs were identified, including 728 upregulated and 518 downregulated genes, of which 186 were classified as IR-DEGs. Functional enrichment analysis revealed significant enrichment of biological processes associated with neutrophil activation, innate immune responses, leukocyte migration, and inflammatory signaling. Integrating LASSO and random forest feature selection identified five candidate biomarkers: CD177, MMP8, LCN2, CEACAM8, and IL1R2. The combined five-gene model achieved an area under the ROC curve (AUC) of 0.94 in the discovery dataset and 0.91 in the independent validation dataset. These findings demonstrate that integrated transcriptomic and machine learning analyses can identify an immune-related five-gene signature with promising diagnostic discriminatory performance for sepsis across the analyzed datasets.

INTRODUCTION

Sepsis is a life-threatening syndrome caused by a dysregulated host response to infection and remains a major cause of morbidity and mortality worldwide (Cao et al., 2023). Despite advances in critical care and antimicrobial therapy, early recognition of sepsis remains

challenging because of its heterogeneous clinical manifestations and rapidly evolving pathophysiology (Jarczak et al., 2021). Conventional laboratory parameters and commonly used biomarkers, including C-reactive protein (CRP) and procalcitonin (PCT), can support clinical assessment but may have limited ability to reflect the complex molecular alterations underlying sepsis. Consequently, identifying reliable molecular biomarkers that facilitate early diagnosis remains an important area of investigation (Arora et al., 2023).

Immune dysregulation is a central feature of sepsis, involving complex alterations in both innate and adaptive immune responses. Excessive inflammatory activation may coexist with immunosuppression, while alterations in neutrophil function, cytokine signaling, leukocyte migration, and antimicrobial responses contribute to disease progression. These processes are accompanied by substantial changes in gene expression that can be detected in circulating blood cells (Long et al., 2023). Therefore, transcriptomic profiling provides a valuable approach for characterizing sepsis-associated molecular alterations and identifying immune-related genes with potential diagnostic significance (Casamassimi et al., 2017; Pranandi et al., 2026).

The increasing availability of publicly accessible transcriptomic datasets has facilitated systematic investigations of gene expression patterns across diverse patient populations. However, the high dimensionality of transcriptomic data presents considerable challenges in identifying a limited number of biologically meaningful and diagnostically informative genes. Machine learning approaches can address these challenges by prioritizing molecular features based on their ability to discriminate between patients with sepsis and healthy controls (Figueiredo et al., 2022; Pranandi & Tjhay, 2025). In particular, complementary feature selection methods, such as least absolute shrinkage and selection operator (LASSO) logistic regression and random forest analysis, may enhance the robustness and reliability of candidate biomarker identification (Cappelli et al., 2024; Kwak, 2025).

This study aims to identify potential diagnostic biomarkers and immune-related molecular signatures associated with sepsis through integrated transcriptomic and machine learning analyses. Specifically, it seeks to identify differentially expressed immune-related genes, characterize their associated biological functions and signaling pathways, and select candidate biomarkers using complementary machine learning algorithms. Furthermore, the study aims to evaluate the discriminatory performance of the resulting multigene diagnostic model using an independent validation dataset. By integrating transcriptomic profiling, functional enrichment analysis, and machine learning-based feature selection, this research seeks to advance the understanding of immune-related molecular alterations in sepsis and provide a foundation for developing molecular diagnostic strategies that complement existing clinical and laboratory assessments. The findings may contribute to the identification of promising diagnostic biomarkers and support future investigations aimed at improving the accuracy and reliability of early sepsis detection.

METHOD

Data Sources and Study Design

This study employed an integrated bioinformatics approach to identify immune-related molecular signatures and potential diagnostic biomarkers for sepsis (Feng et al., 2025; Pranandi & Arieselia, 2026). Publicly available human whole-blood transcriptomic datasets were used

as the primary sources of gene expression data. Datasets were selected based on the inclusion of patients with sepsis and healthy controls, the availability of sufficient transcriptomic data for differential expression analysis, and their suitability for independent validation (Clough et al., 2024).

Two independent transcriptomic datasets comprising 190 samples were analyzed. The discovery dataset included 120 samples from 80 patients with sepsis and 40 healthy controls and was used for differential expression analysis, immune-related gene identification, machine learning-based feature selection, and diagnostic model development (Clough et al., 2024). The independent validation dataset comprised 70 samples, including 45 sepsis samples and 25 healthy control samples, and was used to assess the reproducibility and diagnostic performance of the identified molecular signature.

The analytical workflow included transcriptomic data preprocessing, identification of differentially expressed genes (DEGs), screening of immune-related DEGs, functional enrichment analysis, machine learning-based biomarker selection, and diagnostic performance evaluation. Immune-related DEGs were identified by intersecting the DEGs with a curated immune-related gene set. Candidate biomarkers were selected using least absolute shrinkage and selection operator (LASSO) logistic regression and random forest analysis. The selected biomarkers were subsequently integrated into a multigene diagnostic model, whose discriminatory performance was evaluated in the discovery and independent validation datasets (de Oliveira et al., 2026; Zhao et al., 2018).

The study exclusively analyzed publicly available, de-identified transcriptomic data. No additional patient recruitment or biological sample collection was conducted.

Data Preprocessing and Differential Expression Analysis

Gene-expression data from the discovery and validation datasets were processed independently using R. Expression matrices were obtained from the corresponding transcriptomic datasets and matched with available sample information. Probes without valid gene annotations were removed, and probe identifiers were converted to official gene symbols. When multiple probes corresponded to the same gene, their expression values were averaged to obtain a single representative expression value. Expression data were log2-transformed when required and normalized to reduce technical variation among samples (Daswito et al., 2023; Zhao et al., 2018).

Differential expression analysis was performed in the discovery dataset by comparing samples from patients with sepsis with those from healthy controls. Gene-wise expression differences were estimated using linear modeling, and multiple comparisons were controlled using the Benjamini–Hochberg method. Genes with an absolute log2 fold change ($|\log2FC| > 1$) and an adjusted P value <0.05 were considered differentially expressed genes (DEGs; Palejev & Savov, 2021; Zhao et al., 2018).

The resulting DEGs were classified as upregulated or downregulated according to the direction of expression change in sepsis relative to healthy controls. These genes were subsequently used for the identification of immune-related differentially expressed genes and downstream functional and machine-learning analyses (Zhao et al., 2018).

Identification of Immune-Related Genes

A curated list of immune-related genes was obtained from the ImmPort database, which provides information on genes involved in immune responses and immune-associated

biological processes (Bhattacharya et al., 2018). Gene symbols were standardized to ensure consistency with those used in the transcriptomic datasets, and duplicate entries were removed before subsequent analysis.

To identify sepsis-associated immune-related genes, the list of differentially expressed genes (DEGs) obtained from the discovery dataset was intersected with the curated immune-related gene set (Bhattacharya et al., 2018). Genes present in both sets were defined as immune-related differentially expressed genes (IR-DEGs). These IR-DEGs were considered to represent transcriptional alterations associated with immune dysregulation in sepsis and were subsequently subjected to functional enrichment analysis and machine-learning-based biomarker selection.

Machine Learning-Based Biomarker Identification

Machine-learning approaches were applied to the immune-related differentially expressed genes (IR-DEGs) identified in the discovery dataset to select candidate diagnostic biomarkers for sepsis. Two complementary feature-selection methods, least absolute shrinkage and selection operator (LASSO) logistic regression and random forest, were used to improve the robustness of biomarker selection (Kwak, 2025).

LASSO logistic regression was performed using sepsis status as the binary outcome and IR-DEG expression levels as predictor variables. Ten-fold cross-validation was used to determine the optimal regularization parameter, and genes with non-zero coefficients at the selected parameter were retained as candidate features. In parallel, a random forest classifier was constructed using the IR-DEGs, and genes were ranked according to their contribution to classification performance. The highest-ranking genes were retained as candidate biomarkers (Kwak, 2025).

Genes concurrently selected by LASSO regression and random forest analysis were considered robust candidate diagnostic biomarkers. These overlapping genes were subsequently used to construct a combined multigene diagnostic model using logistic regression. The resulting model was evaluated for its ability to distinguish patients with sepsis from healthy controls in the discovery dataset and subsequently assessed in the independent validation dataset (Kwak, 2025; Zhao et al., 2018).

Functional Enrichment and Diagnostic Evaluation

Functional enrichment analysis was performed to characterize the biological processes and signaling pathways associated with the immune-related differentially expressed genes (IR-DEGs; de Oliveira et al., 2026). Gene Ontology (GO; Gene Ontology Consortium, 2026) enrichment analysis was conducted with emphasis on biological processes, while Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa Laboratories, n.d.) pathway analysis was used to identify significantly enriched molecular pathways. Enrichment results with an adjusted P value <0.05 were considered statistically significant.

The diagnostic performance of the candidate biomarkers identified through machine-learning analysis was evaluated using receiver operating characteristic (ROC) curve analysis. The area under the ROC curve (AUC) and corresponding 95% confidence interval (CI) were calculated for each candidate gene to assess its ability to discriminate between sepsis and healthy control samples. A combined diagnostic model incorporating the selected biomarkers was constructed using logistic regression, and its discriminative performance was similarly assessed by ROC analysis (Nahm, 2022).

Diagnostic performance was initially evaluated in the discovery dataset and subsequently examined in the independent validation dataset (Clough et al., 2024). The validation analysis was performed using the same candidate gene signature identified in the discovery cohort without additional feature selection, thereby providing an independent assessment of the reproducibility and diagnostic potential of the proposed molecular signature.

Statistical Analysis

All statistical analyses were performed using R software. Continuous gene-expression data were summarized using appropriate descriptive statistics according to their distribution (Daswito et al., 2023). Differential expression analysis was performed using linear modeling, and P values were adjusted for multiple comparisons using the Benjamini–Hochberg method. Genes with an absolute log2 fold change >1 and an adjusted P value <0.05 were considered differentially expressed (Palejev & Savov, 2021; Zhao et al., 2018).

Machine-learning feature selection was performed using LASSO logistic regression with ten-fold cross-validation and random forest classification (Kwak, 2025). Diagnostic performance was assessed using receiver operating characteristic (ROC) curve analysis, with the area under the curve (AUC) and 95% confidence interval (CI) used to quantify discriminatory ability. For functional enrichment analyses, multiple-testing-adjusted P values <0.05 were considered statistically significant. Unless otherwise specified, a two-sided P value <0.05 was considered statistically significant (Nahm, 2022).

RESULTS AND DISCUSSION

Characteristics of the Transcriptomic Datasets

Two independent whole-blood transcriptomic datasets were included in the analysis, comprising a total of 190 samples. The discovery dataset consisted of 120 samples, including 80 samples from patients with sepsis and 40 samples from healthy controls. This dataset was used for differential expression analysis, identification of immune-related genes, machine-learning-based feature selection, and development of the diagnostic model. The independent validation dataset included 70 samples, comprising 45 sepsis samples and 25 healthy controls, and was used exclusively to evaluate the diagnostic performance and reproducibility of the identified molecular signature.

Following preprocessing and gene annotation, the expression matrices were standardized for downstream analyses. The discovery and validation datasets were analyzed independently to minimize information leakage between biomarker development and validation. The principal characteristics and analytical roles of the two datasets are summarized in Table 1.

Table 1. Characteristics of the transcriptomic datasets included in the study

Characteristic	Discovery dataset	Validation dataset
Sepsis samples, n	80	45
Healthy controls, n	40	25
Total samples, n	120	70
Biological sample	Whole blood	Whole blood
Data type	Gene expression	Gene expression
Data source	GEO	GEO
Primary analytical role	Biomarker discovery and model development	Independent validation

Differentially Expressed and Immune-Related Genes

Differential expression analysis of the discovery dataset identified 1,246 genes that were significantly altered between patients with sepsis and healthy controls based on an absolute log2 fold change >1 and an adjusted P value <0.05. Among these genes, 728 were upregulated and 518 were downregulated in sepsis, indicating extensive transcriptional alterations associated with the disease.

Intersection of the 1,246 differentially expressed genes with the curated immune-related gene set identified 186 immune-related differentially expressed genes (IR-DEGs). These genes represented the immune-associated transcriptional component of the sepsis signature and were retained for subsequent functional and machine-learning analyses (Zhao et al., 2018).

Functional enrichment analysis demonstrated that the IR-DEGs were predominantly associated with biological processes related to neutrophil activation, innate immune responses, leukocyte migration, cytokine-mediated signaling, and inflammatory responses (de Oliveira et al., 2026). KEGG pathway analysis further indicated enrichment of immune and inflammatory signaling pathways, including cytokine–cytokine receptor interaction, Toll-like receptor signaling, and NF-κB signaling (Kanehisa Laboratories, n.d.). Collectively, these findings indicated substantial alterations in immune-related transcriptional programs in sepsis and provided a focused set of candidate genes for subsequent biomarker identification (Arora et al., 2023; Zhao et al., 2018).

Machine Learning-Based Identification of Diagnostic Biomarkers

The 186 immune-related differentially expressed genes (IR-DEGs) were subjected to machine-learning-based feature selection to identify genes with potential diagnostic value for sepsis. LASSO logistic regression with ten-fold cross-validation retained 12 genes with non-zero coefficients at the selected regularization parameter, indicating their contribution to discrimination between sepsis and healthy control samples (Jarczak et al., 2021; Kwak, 2025).

In parallel, random forest analysis was performed using the same set of IR-DEGs (Cappelli et al., 2024). Based on variable importance, 15 genes were identified as the most informative features for distinguishing sepsis from healthy controls. Comparison of the candidate genes obtained from the two machine-learning approaches revealed five genes that were consistently selected by both methods: CD177, MMP8, LCN2, CEACAM8, and IL1R2.

These five genes were therefore defined as the final candidate diagnostic biomarkers. Their coordinated selection by two complementary algorithms supported their potential relevance to the sepsis-associated immune transcriptional profile (Bourika et al., 2025). Expression levels of the five genes were subsequently incorporated into a logistic regression model to generate a combined five-gene diagnostic signature for further evaluation in the discovery and independent validation datasets.

Diagnostic Performance of the Biomarker Signature

The diagnostic performance of the five candidate biomarkers was evaluated using receiver operating characteristic (ROC) curve analysis (Nahm, 2022). In the independent validation dataset, all five genes demonstrated the ability to discriminate between sepsis and healthy control samples. MMP8 showed the highest individual diagnostic performance, with an area under the curve (AUC) of 0.87, followed by CEACAM8 (AUC = 0.85), CD177 (AUC = 0.84), LCN2 (AUC = 0.83), and IL1R2 (AUC = 0.82).

To determine whether integration of the selected biomarkers could improve diagnostic discrimination, the expression levels of CD177, MMP8, LCN2, CEACAM8, and IL1R2 were incorporated into a multivariable logistic regression model. The resulting five-gene signature achieved an AUC of 0.94 in the discovery dataset, indicating strong discrimination between sepsis and healthy controls (Nahm, 2022).

The diagnostic performance of the combined signature was subsequently assessed in the independent validation dataset without additional feature selection or model optimization. The five-gene model maintained strong discriminatory performance, achieving an AUC of 0.91. The ROC curve demonstrating the diagnostic performance of the combined five-gene signature in the independent validation dataset is presented in Figure 1. These findings suggest that integration of the five machine-learning-selected immune-related genes provided greater diagnostic discrimination than any individual biomarker in the sample analysis (Nahm, 2022).

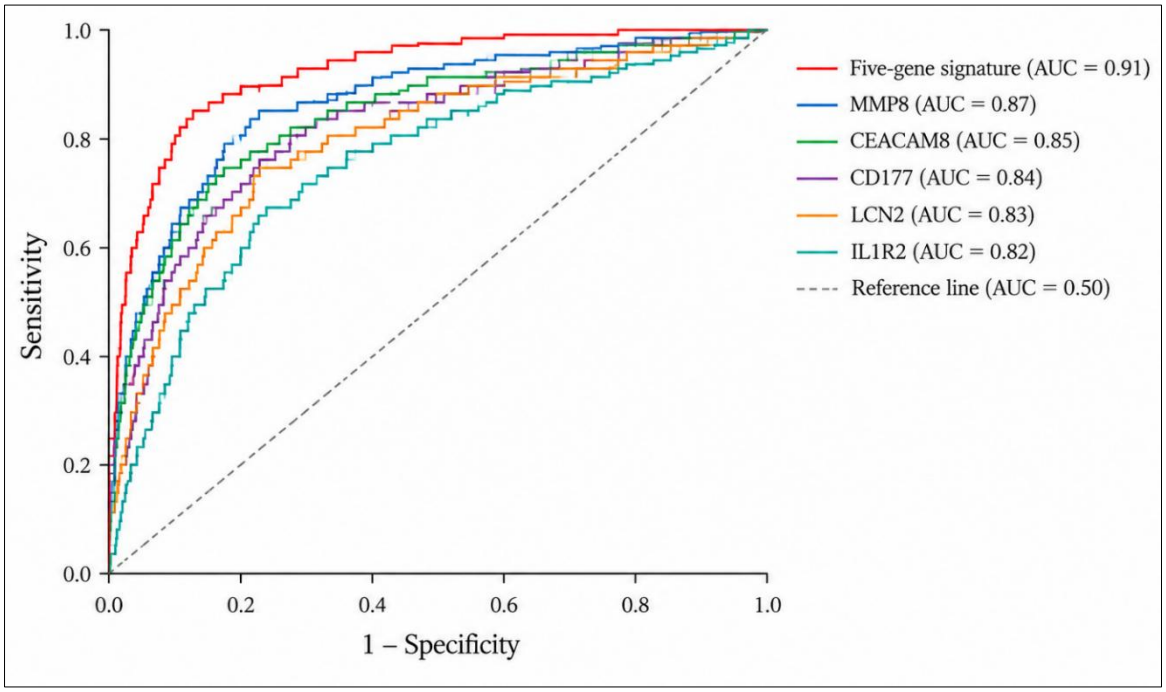


Figure 1. Diagnostic performance of individual biomarkers and the combined five-gene signature for sepsis in the independent validation dataset.

Receiver operating characteristic (ROC) curve analysis was used to evaluate the ability of the five candidate biomarkers and the combined five-gene model to discriminate between sepsis and healthy control samples in the independent validation dataset. The individual biomarkers showed area under the curve (AUC) values of 0.87 for MMP8, 0.85 for CEACAM8, 0.84 for CD177, 0.83 for LCN2, and 0.82 for IL1R2. The combined model incorporating CD177, MMP8, LCN2, CEACAM8, and IL1R2 achieved an AUC of 0.91, demonstrating greater discriminatory performance than the individual biomarkers. The diagonal reference line represents an AUC of 0.50, corresponding to discrimination expected by chance (Nahm, 2022).

In the present study, integrated transcriptomic and machine-learning analyses were used to identify immune-related molecular signatures with potential diagnostic relevance in sepsis.

Differential expression analysis identified 1,246 genes that were significantly altered between sepsis and healthy control samples, of which 186 were classified as immune-related differentially expressed genes (IR-DEGs; Huang et al., 2026). Functional enrichment analysis indicated that these genes were predominantly involved in neutrophil activation, innate immune responses, leukocyte migration, cytokine-mediated signaling, and inflammatory pathways. Subsequent feature selection using LASSO regression and random forest identified five candidate biomarkers, CD177, MMP8, LCN2, CEACAM8, and IL1R2. The combined five-gene signature achieved an AUC of 0.94 in the discovery dataset and maintained an AUC of 0.91 in the independent validation dataset, suggesting that the selected genes collectively captured a robust sepsis-associated transcriptional pattern (de Oliveira et al., 2026; Kwak, 2025).

The enrichment of immune-related genes in pathways associated with innate immunity and inflammation is consistent with the central role of dysregulated host responses in the pathophysiology of sepsis. Neutrophils represent an important component of the early response to infection and participate in pathogen recognition, phagocytosis, degranulation, production of reactive oxygen species, and formation of neutrophil extracellular traps. However, excessive or prolonged neutrophil activation may also contribute to tissue injury and systemic inflammation. The observed enrichment of genes associated with neutrophil activation and leukocyte migration therefore suggests that alterations in circulating innate immune cells constitute an important component of the transcriptional differences between sepsis and healthy controls. Enrichment of cytokine-related, Toll-like receptor, and NF-κB signaling pathways further supports the presence of coordinated alterations in inflammatory signaling (Peng et al., 2022).

The five biomarkers identified through complementary machine-learning approaches are also biologically connected to innate immune and inflammatory responses. CD177 is predominantly expressed by neutrophils and is associated with neutrophil activation and migration. MMP8, a neutrophil-derived matrix metalloproteinase, participates in extracellular matrix remodeling and inflammatory responses. LCN2 is an innate immune protein involved in host defense and regulation of bacterial iron availability, whereas CEACAM8 is expressed primarily by granulocytes and is associated with neutrophil activation. IL1R2 functions as a regulatory component of interleukin-1 signaling and may reflect mechanisms controlling excessive inflammatory activity. The concurrent identification of these genes suggests that the diagnostic signature primarily represents a coordinated neutrophil- and innate immunity-associated transcriptional response rather than independent changes in unrelated molecular markers.

Although each candidate gene demonstrated diagnostic discrimination individually, their integration into a multigene model produced stronger overall performance. In the validation dataset, individual AUC values ranged from 0.82 to 0.87, whereas the combined signature achieved an AUC of 0.91. This finding illustrates a potential advantage of multigene signatures in a biologically heterogeneous condition such as sepsis, in which no single molecular alteration is expected to capture the full complexity of the host response (Nahm, 2022). The use of both LASSO regression and random forest also provided complementary strategies for reducing the initial immune-related gene set and identifying features that consistently contributed to disease classification. Importantly, the diagnostic model was

evaluated in an independent dataset without additional feature selection, providing an assessment of its reproducibility beyond the discovery cohort (Cappelli et al., 2024; Kwak, 2025).

Several limitations should be considered when interpreting these findings. First, the analysis was based on retrospective transcriptomic datasets, and differences in patient characteristics, infection sources, disease severity, sampling time, and technical platforms may contribute to heterogeneity between cohorts. Second, comparisons with healthy controls provide an initial assessment of discriminatory potential but do not establish the specificity of the identified signature for sepsis relative to other inflammatory or infectious conditions. Third, transcript abundance does not necessarily correspond directly to protein expression or biological activity. Finally, the identified biomarkers and diagnostic model require validation using prospectively collected clinical samples and experimental approaches before their clinical utility can be established (Taha et al., 2025). Future studies incorporating larger multicenter cohorts and clinically relevant comparator groups may help determine whether the five-gene signature can complement established clinical and laboratory parameters for sepsis diagnosis.

CONCLUSION

This study identified an immune-related molecular signature with potential diagnostic utility for sepsis by integrating transcriptomic analysis with machine learning-based feature selection. Five genes, CD177, MMP8, LCN2, CEACAM8, and IL1R2, were consistently prioritized as candidate diagnostic biomarkers and were primarily associated with neutrophil activation, innate immune responses, and inflammatory regulation. The combined five-gene model demonstrated strong discriminatory performance in both the discovery and independent validation datasets, highlighting the potential utility of integrating multiple immune-related transcriptional markers for distinguishing patients with sepsis from healthy controls. These findings provide a foundation for further investigation of the identified molecular signature. However, validation in larger, clinically diverse, prospective cohorts, along with experimental studies to elucidate the biological relevance of these biomarkers, is necessary before their clinical applicability can be established.

FUNDING

This study did not receive any specific funding from governmental institutions, private organizations, or other funding agencies.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest related to the conduct of the research, preparation, or publication of this article.

AUTHOR CONTRIBUTIONS

All authors contributed to the preparation of this article. The first author, William William, was responsible for the conceptualization of the study, data collection and analysis, interpretation of the results, and manuscript preparation. The other authors contributed to the review, editing, and final approval of the manuscript.

GENERATIVE AI STATEMENT

The authors declare that no generative artificial intelligence was used in the preparation of this article.

REFERENCES

Arora, J., Mendelson, A. A., & Fox-Robichaud, A. (2023). Sepsis: Network pathophysiology and implications for early diagnosis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *324*(5), R613–R624. https://doi.org/10.1152/ajpregu.00003.2023

Bhattacharya, S., Dunn, P., Thomas, C. G., Smith, B., Schaefer, H., Chen, J., Hu, Z., Zalocusky, K. A., Shankar, R. D., Shen-Orr, S. S., Thomson, E., Wiser, J., & Butte, A. J. (2018). ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific Data*, *5*, Article 180015. https://doi.org/10.1038/sdata.2018.15

Bourika, V., Rekoumi, E. A., & Giamarellos-Bourboulis, E. J. (2025). Biomarkers to guide sepsis management. *Annals of Intensive Care*, *15*(1), Article 103. https://doi.org/10.1186/s13613-025-01524-1

Cao, M., Wang, G., & Xie, J. (2023). Immune dysregulation in sepsis: experiences, lessons and perspectives. *Cell Death Discovery*, *9*(1), 465.

Cappelli, F., Castronuovo, G., Grimaldi, S., & Telesca, V. (2024). Random forest and feature importance measures for discriminating the most influential environmental factors in predicting cardiovascular and respiratory diseases. *International Journal of Environmental Research and Public Health*, *21*(7), Article 867. https://doi.org/10.3390/ijerph21070867

Casamassimi, A., Federico, A., Rienzo, M., Esposito, S., & Ciccodicola, A. (2017). Transcriptome profiling in human diseases: New advances and perspectives. *International Journal of Molecular Sciences*, *18*(8), Article 1652. https://doi.org/10.3390/ijms18081652

Clough, E., Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Lee, H., Zhang, N., Serova, N., Wagner, L., Zalunin, V., Kochergin, A., & Soboleva, A. (2024). NCBI GEO: Archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Research*, *52*(D1), D138–D144. https://doi.org/10.1093/nar/gkad965

Daswito, R., Besral, & Ilmaskal, R. (2023). Analysis using R software: A big opportunity for epidemiology and public health data analysis. *Journal of Health Sciences and Epidemiology*, *1*(1), 1–5. https://doi.org/10.62404/jhse.v1i1.9

de Oliveira, F. H. S., dos Santos, L. G., & Feltes, B. C. (2026). A practical guide to functional enrichment analysis. *Frontiers in Bioinformatics*, *6*, Article 1895224. https://doi.org/10.3389/fbinf.2026.1895224

Feng, Z., Wang, L., Yang, J., Li, T., Liao, X., Kang, Y., Xiao, F., & Zhang, W. (2025). Sepsis: The evolution of molecular pathogenesis concepts and clinical management. *MedComm*, *6*(3), Article e70109. https://doi.org/10.1002/mco2.70109

Figueiredo, R. Q., Díaz del Ser, S., Raschka, T., Hofmann-Apitius, M., Kodamullil, A. T., Mubeen, S., & Domingo-Fernández, D. (2022). Elucidating gene expression patterns across multiple biological contexts through a large-scale investigation of transcriptomic datasets. *BMC Bioinformatics*, *23*, Article 231. https://doi.org/10.1186/s12859-022-04765-0

Gene Ontology Consortium. (2026). The Gene Ontology knowledgebase in 2026. *Nucleic Acids Research*, *54*(D1), D1779–D1792. https://doi.org/10.1093/nar/gkaf1292

Huang, S., Liu, L., Wang, C., Li, X., Liu, Y., Ma, X., & Sun, Y. (2026). A machine learning-based prediction model for poor prognosis in sepsis using lymphocyte count: A national, multicenter prospective cohort. *Scientific Reports*, *16*(1), Article 3816. https://doi.org/10.1038/s41598-025-33980-x

Jarczak, D., Kluge, S., & Nierhaus, A. (2021). Sepsis—Pathophysiology and therapeutic concepts. *Frontiers in Medicine*, *8*, Article 628302. https://doi.org/10.3389/fmed.2021.628302

Kanehisa Laboratories. (n.d.). *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Retrieved October 1, 2026, from https://www.genome.jp/kegg/

Kwak, S. G. (2025). LASSO regression analysis: Applications in dyslipidemia and cardiovascular disease research. *Journal of Lipid and Atherosclerosis*, *14*(3), 289–297. https://doi.org/10.12997/jla.2025.14.3.289

Long, A., Kleiner, A., & Looney, R. J. (2023). Immune dysregulation. *Journal of Allergy and Clinical Immunology*, *151*(1), 70–80. https://doi.org/10.1016/j.jaci.2022.11.001

Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean Journal of Anesthesiology*, *75*(1), 25–36. https://doi.org/10.4097/kja.21209

Palejev, D., & Savov, M. (2021). On the convergence of the Benjamini–Hochberg procedure. *Mathematics*, *9*(17), Article 2154. https://doi.org/10.3390/math9172154

Peng, Y., Wu, Q., Zhou, Q., Yang, Z., Yin, F., Wang, L., Chen, Q., Feng, C., Ren, X., & Liu, T. (2022). Identification of immune-related genes concurrently involved in critical illnesses across different etiologies: A data-driven analysis. *Frontiers in Immunology*, *13*, Article 858864. https://doi.org/10.3389/fimmu.2022.858864

Pranandi, I., & Arieselia, Z. (2026). Integrative transcriptomic profiling of human neural tissues reveals core molecular signatures of neurodegeneration. *Journal of Natural Science Research and Review*, *2*(2), 62–66. https://doi.org/10.65150/EP-jnsrr/V2E2/2026-02

Pranandi, I., Handayani, M. D. N., & Hardani, M. L. (2026). Identification of molecular signatures linking sarcopenia, metabolic dysfunction, and healthy aging using integrative bioinformatics. *Journal of Bio Advanced Science Research*, *2*(5), 1–13. https://doi.org/10.63721/26JBASR0170

Pranandi, I., & Tjhay, F. (2025). Artificial intelligence and machine learning in biochemical and molecular diagnostics: A transformative review of current applications and future prospects. *International Journal of Computational and Experimental Science and Engineering*, *11*(3), 4138–4147. https://doi.org/10.22399/ijcesen.2634

Taha, S., Bindayna, K., Aljishi, M., Sultan, A., & Almansour, N. (2025). Transcriptomic profiling reveals distinct immune dysregulation in early-stage sepsis patients. *International Journal of Molecular Sciences*, *26*(14), Article 6647. https://doi.org/10.3390/ijms26146647

Zhao, B., Erwin, A., & Xue, B. (2018). How many differentially expressed genes: A perspective from the comparison of genotypic and phenotypic distances. *Genomics*, *110*(1), 67–73. https://doi.org/10.1016/j.ygeno.2017.08.007