Predicting Mortality and the Need for Early Intubation in Intensive Care Patients Using Machine Learning: A Pilot Study

Seval KILBASANLI[*], Andaç Batur ÇOLAK[**], Şerife Buket BOZKURT POLAT[***], Zeynep YÜKSEL TURHAN[****], Mustafa KAÇMAZ[*****], Seyyid Mehmet BULUT[******]

Abstract

Aim: This pilot study aimed to assess the feasibility and preliminary performance of machine learning models for predicting ICU mortality and the need for intubation within 48 hours using routinely collected admission data.

Method: Ten adult intensive care patients were included in this single-center, retrospective observational pilot study. For ANN development, 13 prespecified clinical and laboratory features per patient were restructured in long format, yielding 130 feature-level records. These records represented feature entries derived from 10 patients rather than independent patient-level observations. Demographic characteristics, primary diagnosis, APACHE II and SOFA scores, arterial blood gas parameters, oxygenation indicators (FiO_2, PaO_2/FiO_2), and routine metabolic–biochemical variables were analyzed. ML-based classification models predicted ICU mortality and the need for intubation within the first 48 hours.

Results: In the exploratory feature-level internal assessment, both the ICU mortality and early-intubation classifications yielded an AUC of 1.00, sensitivity of 100%, specificity of 100%, and an F1-score of 1.00. These estimates were derived from records originating from 10 patients and should be interpreted cautiously.

Conclusion: This pilot study demonstrates the feasibility of applying machine learning methods to routinely collected ICU admission data. The findings are preliminary and hypothesis-generating and should not be interpreted as established clinical predictive performance. Validation in larger, independent, prospective, and multicenter cohorts is required.

Özgün Araştırma Makalesi (Original Research Article)
Geliş / Received: 02.02.2026 **Kabul / Accepted:** 20.07.2026
DOI: https://doi.org/10.38079/igusabder.1880038

[*] (Corresponding Author) Asst. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Anesthesiology and Reanimation, Niğde, Türkiye. E-mail: sevalkilbasanli@gmail.com ORCID https://orcid.org/0000-0003-4052-9593
[**] Assoc. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Computer and Information Sciences, Department of Information Systems and Technologies, Niğde, Türkiye. E-mail: bcolak@ohu.edu.tr ORCID https://orcid.org/0000-0001-9297-8134
[***] Assoc. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Medical Biochemistry, Niğde, Türkiye. E-mail: buketbozkurt@yahoo.com ORCID https://orcid.org/0000-0002-8641-2844
[****] Asst. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Anesthesiology and Reanimation, Niğde, Türkiye. E-mail: zeynepyuksel6@gmail.com ORCID https://orcid.org/0000-0002-8906-2149
[*****] Assoc. Prof. Dr., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Anesthesiology and Reanimation, Niğde, Türkiye. E-mail: muskac51@gmail.com ORCID https://orcid.org/0000-0002-8655-3882
[******] Res. Asst., Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Medical Biochemistry, Niğde, Türkiye. E-mail: sbulut@ohu.edu.tr ORCID https://orcid.org/0000-0002-2010-7230

ETHICAL STATEMENT: The protocol was approved by the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University, Faculty of Medicine, Türkiye (Decision No: 2026/3, Date: 08.01.2026).

Keywords: Intensive care, machine learning, mortality, intubation, pilot study, decision support.

Yoğun Bakım Hastalarında Mortalite ve Erken Entübasyon Gereksiniminin Makine Öğrenmesi ile Öngörülmesi: Pilot Çalışma

Öz

Amaç: Bu pilot çalışmada, rutin yatış verilerine dayalı makine öğrenmesi modellerinin yoğun bakım mortalitesi ile ilk 48 saat içinde entübasyon gereksinimini öngörmedeki uygulanabilirliğinin ve ön performansının değerlendirilmesi amaçlandı.

Yöntem: Tek merkezli, retrospektif, gözlemsel pilot çalışmaya 10 erişkin yoğun bakım hastası dahil edildi. Yapay sinir ağı geliştirilmesi için her hastaya ait önceden belirlenmiş 13 klinik ve laboratuvar özelliği uzun veri formatında yeniden düzenlenerek toplam 130 özellik düzeyinde kayıt oluşturuldu. Bu kayıtlar, bağımsız hasta gözlemlerini değil, 10 hastadan türetilen özellik kayıtlarını temsil etmekteydi. Demografik özellikler, primer tanı, APACHE II ve SOFA skorları, arter kan gazı parametreleri, oksijenasyon göstergeleri (FiO_2, PaO_2/FiO_2) ve rutin metabolik–biyokimyasal değişkenler analiz edildi. Yoğun bakım mortalitesi ve ilk 48 saat içinde entübasyon gereksiniminin öngörülmesi amacıyla ML tabanlı sınıflandırma modelleri kullanıldı.
Bulgular: Özellik düzeyindeki keşifsel iç değerlendirmede, hem yoğun bakım mortalitesi hem de erken entübasyon sınıflandırması için AUC 1,00, duyarlılık %100, özgüllük %100 ve F1-skoru 1,00 olarak bulundu. Bu tahminler 10 hastadan türetilen kayıtlara dayandığından dikkatle yorumlanmalıdır.

Sonuç: Bu pilot çalışma, yoğun bakım yatışında rutin olarak elde edilen verilere makine öğrenmesi yöntemlerinin uygulanabilir olduğunu göstermektedir. Bulgular ön ve hipotez oluşturucu nitelikte olup, yerleşik bir klinik öngörü performansının kanıtı olarak değerlendirilmemelidir. Modellerin daha geniş, bağımsız, prospektif ve çok merkezli hasta gruplarında doğrulanması gereklidir.

Anahtar Sözcükler: Yoğun bakım, makine öğrenmesi, mortalite, entübasyon, pilot çalışma, karar destek sistemi.

Introduction

Intensive care units are clinical settings where life-threatening acute illnesses and multiple organ failures are managed and where the risk of mortality is high. Predicting mortality risk early on is valuable for planning the intensity of monitoring, timing interventions, and using resources effectively. In recent years, numerous studies and reviews have been published showing that machine learning (ML) approaches can be effective in predicting intensive care mortality[1–6].

The need for intubation is one of the most critical indicators of clinical deterioration in intensive care patients. Delayed intubation may be associated with difficulties in airway management, increased risk of complications, and poor clinical outcomes. Recently, externally validated studies and deep learning-based approaches have been reported to predict the need for intubation using interpretable ML models; models targeting time-dependent outcomes, such as the time to intubation during the first days of intensive care, have been developed[7–9].

Sepsis remains one of the most important predictors of intensive care mortality. Physiological deterioration in sepsis can progress rapidly, making early risk classification difficult. Approaches addressing imbalanced class problems have been shown to improve sepsis mortality prediction. Machine learning models have also been developed to predict mortality in patients with sepsis and in those with concomitant chronic kidney disease[10-12].

Although scores such as APACHE II and SOFA are widely used in clinical practice for mortality prediction, these scores have limitations in terms of early and accurate prediction at the individual level. In contrast, it has been reported that easily accessible variables such as electrolyte imbalances and routine laboratory parameters can emerge as meaningful predictors in ML-based models[13].

Intensive care patients on mechanical ventilation have a higher risk of mortality. ML models predicting hospital mortality and community approaches predicting prolonged ventilation requirements have been reported in this group. Studies analyzing factors associated with invasive mechanical ventilation using ML and models focusing on early mortality prediction in ARDS reveal significant potential for clinical decision support systems. A recent overview of artificial intelligence applications in mechanical ventilation emphasizes that decision support systems working with routine clinical data will become increasingly prominent[14-18].

In this context, alongside models developed with large datasets, it is important that small-sample pilot studies conducted with variables accessible in routine clinical practice demonstrate feasibility and lay the methodological groundwork for larger studies. Furthermore, it has been reported that not only mortality but also outcomes related to resource utilization, such as length of stay, can be predicted using ML methods in intensive care admissions[2]. In this pilot study, we aimed to evaluate the predictability of intensive care mortality and the need for intubation within the first 48 hours using ML methods based on routine data obtained at the time of admission to the intensive care unit.

Material and Methods

Study Design and Participants

This was a single-center, retrospective, observational pilot study. Adult patients aged 18 years or older who were admitted to the intensive care unit between December 1 and 31, 2025, were retrospectively screened. Patients with complete data for the prespecified admission variables and study outcomes were eligible for inclusion. Patients younger than 18 years and those with missing clinical, laboratory, or outcome data were excluded. Eligible patients were included consecutively according to the order of ICU admission.

Predictor Variables and Outcomes

Predictor variables were obtained at ICU admission or within the first few hours and included age, sex, primary diagnosis, APACHE II and SOFA scores, arterial blood gas parameters (pH, PaO_2, $PaCO_2$, and HCO_3^-), lactate, glucose, sodium, potassium, calcium, FiO_2, and the PaO_2/FiO_2 ratio. The two prespecified outcomes were ICU mortality and new endotracheal intubation within 48 hours after ICU admission. ICU mortality was evaluated in the entire cohort. The early-intubation outcome was evaluated only among patients who were not intubated at ICU admission. Patients already intubated at admission remained included in the mortality analysis but were excluded from the early-intubation analysis. Baseline intubation status was not used as a predictor in the early-intubation model.

Artificial Neural Network Development

The study included 10 unique patients. For ANN input construction, 13 prespecified clinical and laboratory features from each patient were restructured in long format, producing 130 feature-level records (10 patients×13 features). Each record included the corresponding patient-level demographic and diagnostic information, a coded clinical or laboratory feature, and the value of that feature; ICU mortality and intubation within 48 hours were specified as outcome labels. Accordingly, the 130 records represented feature-level entries derived from 10 patients and did not constitute 130 independent patients or independent clinical observations. This study employed a multilayer perceptron (MLP) architecture to train an ANN for the prediction of potentially detrimental clinical outcomes. We selected this feed-forward network architecture due to its capacity to characterize non-linear, intricate interactions between prevalent admission factors and life-threatening patient occurrences. The Levenberg-Marquardt algorithm, a powerful tool for improving MLP framework weights, was used to train the model.

The predictor inputs included age, sex, primary diagnosis, APACHE II and SOFA scores, arterial blood gas parameters, lactate, glucose, electrolytes, FiO_2, and the PaO_2/FiO_2 ratio. Baseline intubation status was included only in the ICU mortality model. For the early-intubation model, only patients who were not intubated at ICU admission were included, and baseline intubation status was excluded from the predictor set. Separate classification outputs were generated for ICU mortality and the need for new endotracheal intubation within 48 hours.

The 130 feature-level records were allocated to internal model-development subsets comprising 90 records for training, 20 for validation, and 20 for testing. This partitioning was used for exploratory internal model assessment and did not constitute independent patient-level validation. The MLP architecture included a single hidden layer with 20 neurons.

Model performance was evaluated using mean squared error (MSE), the correlation coefficient (R), margin of deviation (MoD), area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and F1-score. AUC was calculated from the continuous ANN output scores as the area under the receiver operating characteristic curve. Sensitivity, specificity, and F1-score were calculated from the corresponding binary classification results. These metrics were used for exploratory internal assessment of the feature-level training, validation, and testing subsets and were not considered measures of independent patient-level validation. The mathematical expressions for MSE, R, MoD, sensitivity, specificity, and F1-score are provided below.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (X_{targ\,(i)} - X_{pred(i)})^2 \tag{1}$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^{N} \left(X_{(i)} - X_{pred(i)}\right)^2}{\sum_{i=1}^{N} \left(X_{(i)}\right)^2}} \tag{2}$$

$$MoD\ (\%) = \left[\frac{X_{targ} - X_{pred}}{X_{targ}}\right] x\ 100 \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{6}$$

In these equations, X_{targ} represents the actual clinical values, X_{pred} denotes the values predicted by the ANN model, while TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives obtained from the classification confusion matrix, respectively.

Ethical Statement

The protocol was approved by the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University Faculty of Medicine, Türkiye (Decision No: 2026/3, Date: 08.01.2026)

Results

Figure 1 presents the MSE values across 12 training epochs. At epoch 12, the MSE values were 5.46×10^{-15} for the training subset, 9.95×10^{-15} for the validation subset, and 5.63×10^{-15} for the testing subset. In the exploratory feature-level internal assessment, both the ICU mortality and early-intubation classifications yielded an AUC of 1.00, sensitivity of 100%, specificity of 100%, and an F1-score of 1.00.

Figure 1. The performance of the developed ANN model during its development phase

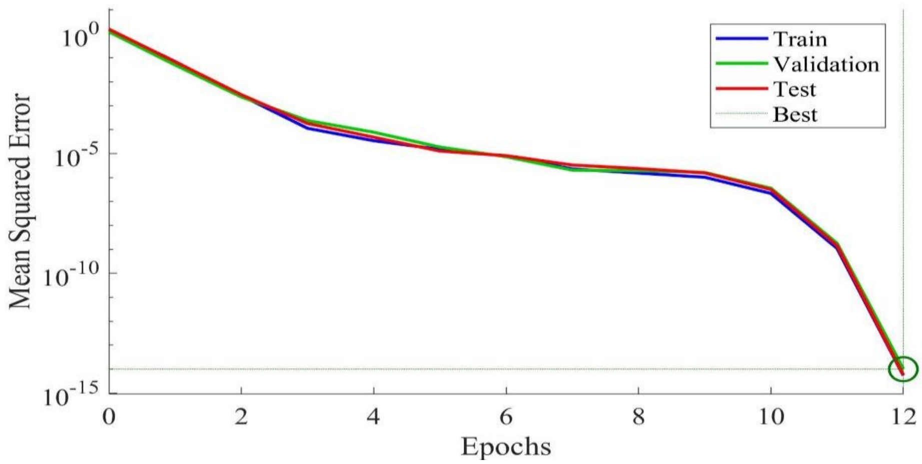
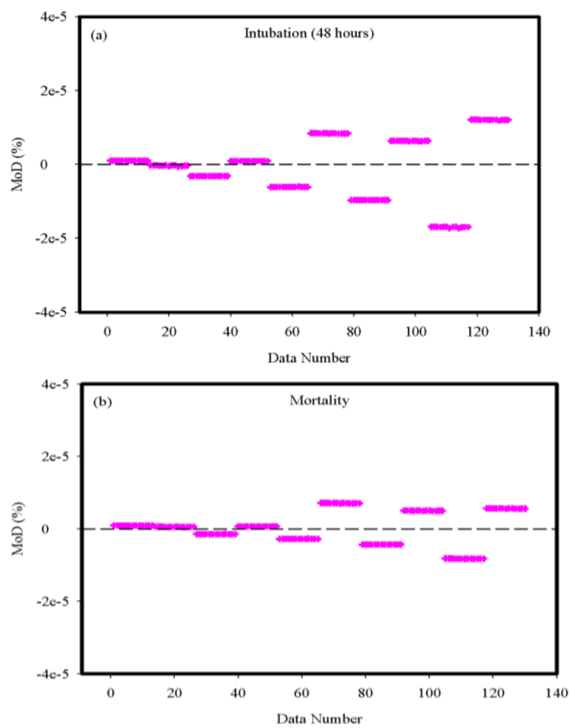


Figure 2 presents the MoD values across the feature-level records used for the internal assessment of each outcome. The deviations ranged approximately from -4×10^{-5} to $+4\times10^{-5}$. The mean MoD was -7.39×10^{-7}% for intubation within 48 hours and 3.32×10^{-7}% for ICU mortality.

Figure 2. Distribution of the MoD across the feature-level records used for the internal assessment of (a) intubation within 48 hours and (b) ICU mortality.



The R value was 0.99999 for the training, validation, and testing subsets.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Discussion

This pilot study provides preliminary evidence that an artificial neural network framework can be applied to routinely collected ICU admission data to explore mortality and the need for early intubation. However, considering the inclusion of only 10 patients and the high model complexity relative to the patient sample, the near-perfect internal performance estimates should be interpreted primarily as evidence of methodological feasibility rather than stable or generalizable predictive accuracy. These findings are exploratory and intended to guide the design of larger validation studies. Within this context, the overall direction of our findings is consistent with previous mortality prediction studies conducted using substantially larger datasets[1–6].

Among the predictor domains incorporated into the ANN framework, illness-severity scores, oxygenation measures, and metabolic variables have clear clinical relevance to the outcomes examined. APACHE II and SOFA scores reflect the extent of acute physiological disturbance and multisystem organ dysfunction and may therefore contain information relevant to both mortality and respiratory deterioration. Oxygenation variables, particularly FiO_2 and the PaO_2/FiO_2 ratio, reflect the severity of gas-exchange impairment and are clinically relevant to the potential need for early intubation. Similarly, elevated lactate and abnormalities in acid–base status, glucose, and electrolytes may indicate tissue hypoperfusion, metabolic stress, and loss of physiological homeostasis. However, because no separate formal feature-importance analysis was performed and the patient sample was very small, these observations should be regarded as clinically guided, hypothesis-generating interpretations rather than evidence of a definitive predictor ranking or independent prognostic effects. The clinical relevance of these predictor domains is consistent with previous machine learning studies of mortality and intubation outcomes in critically ill patients[7–13,19].

Early risk classification in patients with sepsis is challenging due to heterogeneous clinical courses. Methods addressing imbalanced data have been shown to improve sepsis mortality prediction, and machine learning models have also been developed for mortality prediction in patients with sepsis and in those with concomitant chronic kidney disease[10-12]. In this context, the hypothesis-generating value of pilot studies is important.

The risk of mortality and adverse clinical outcomes is substantially higher among critically ill patients requiring mechanical ventilation. Recent machine learning–based models have demonstrated the ability to predict hospital mortality in mechanically ventilated intensive care patients, highlighting the prognostic importance of early respiratory deterioration. Community- and cohort-based approaches predicting prolonged mechanical ventilation requirements further emphasize the clinical and resource-related implications of timely risk stratification. Studies analyzing factors associated with invasive mechanical ventilation using machine learning techniques, as well as models focusing on early mortality prediction in acute respiratory distress syndrome (ARDS), illustrate the evolving role of data-driven decision support systems in intensive care. A recent overview also underscores that artificial intelligence–based

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

applications may increasingly contribute to the management of mechanical ventilation in critical care practice[14–18].

One of the strengths of the present study is that all variables used in model development are readily available in routine clinical practice at the time of intensive care admission. Demonstrating that commonly encountered parameters, such as electrolyte imbalances and standard laboratory values, can meaningfully contribute to machine learning–based mortality prediction enhances the clinical applicability of such models. This observation is consistent with previous reports indicating that routinely measured biochemical variables may provide substantial predictive value when incorporated into machine learning frameworks[13].

Limitations

This study has several important limitations. First, the inclusion of only 10 patients substantially limits the statistical stability, representativeness, and generalizability of the findings. The high model complexity relative to the number of individual patients creates a substantial risk of overfitting and may contribute to the near-perfect internal performance estimates. Moreover, internal partitioning into training, validation, and testing subsets cannot substitute for validation in an independent patient cohort. In addition, the 130 feature-level records were derived from only 10 patients and were not statistically independent. Because partitioning was performed at the feature-record level, records originating from the same patient may have been represented in more than one model-development subset, creating a risk of information leakage and overly optimistic internal performance estimates. Therefore, the reported findings should be interpreted as exploratory evidence of methodological feasibility rather than as evidence of established clinical predictive accuracy. The retrospective, single-center design further limits external validity. Future studies should include substantially larger, prospectively collected, multicenter cohorts and perform independent external validation with patient-level separation across all model-development subsets.

Conclusions

This pilot study explored the methodological feasibility of applying an artificial neural network model trained on routinely collected ICU admission data to assess ICU mortality and the need for intubation within 48 hours. The developed multilayer perceptron model, which included 20 hidden neurons, yielded an R value of 0.99999 and very low mean squared error values across the internal training, validation, and testing subsets. However, given the inclusion of only 10 patients and the high model complexity relative to the patient sample, these near-perfect internal performance estimates may reflect overfitting and should not be interpreted as evidence of established clinical accuracy or generalizability. Routine admission variables may contain potentially informative signals for early risk assessment; however, this observation requires confirmation in substantially larger, prospective, multicenter cohorts with independent external validation. Therefore, the present findings should be considered preliminary and hypothesis-generating rather than evidence of a clinically applicable predictive model.

Conflict of Interest Statement: The authors declare no conflict of interest related to this study.

Financial Disclosure: The authors declare that no financial support was received during the conduct of this study.

Ethical Approval: Ethical approval for this study was obtained from the Non-Interventional Clinical Research Ethics Committee of Niğde Ömer Halisdemir University Faculty of Medicine (Decision No: 2026/3). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Author Contributions

SK: Conceptualization, study design, data collection, clinical interpretation, supervision, and writing. **ABÇ:** Writing - original draft, methodology, software, conceptualization, and formal analysis. **ŞBBP:** Conceptualization, study supervision, and critical review of the manuscript. **ZYT, MK:** Clinical patient care and general supervision. **SMB:** Data support

REFERENCES

1. Sun H, Kang M, Zhang H, Jia J, Wang Q. Machine learning for predicting mortality in intensive care unit patients: a prognostic performance systematic review and meta-analysis. *Nursing in Critical Care*. 2025;30(6):e70206.

2. Iwase S, Nakada T, Shimada T, et al. Prediction algorithm for ICU mortality and length of stay using machine learning. *Scientific Reports*. 2022;12:12912.

3. Pang K, Li L, Ouyang W, Liu X, Tang Y. Establishment of ICU mortality risk prediction models with machine learning algorithm using MIMIC-IV database. *Diagnostics*. 2022;12(5):1068.

4. Lim L, Gim U, Cho K, et al. Real-time machine learning model to predict short-term mortality in critically ill patients: development and international validation. *Critical Care*. 2024;28(1):76.

5. Thadani S, Wu TC, Wu DTY, et al. Machine learning-based prediction model for ICU mortality after continuous renal replacement therapy initiation in children. *Critical Care Explorations*. 2024;6(12):e1188.

6. Yang S, Sun Y, Wang T, et al. Machine learning-based prediction of mortality and multidrug-resistant infection risks in ICU patients with suspected infection: a prospective national multicenter cohort study. *BMC Infectious Diseases*. 2026;26:139.

7. Liu J, Duan X, Duan M, et al. Development and external validation of an interpretable machine learning model for the prediction of intubation in the intensive care unit. *Scientific Reports*. 2024;14:27174.

8. Li R, Xu Z, Xu J, et al. Predicting intubation for intensive care units patients: a deep learning approach to improve patient management. *International Journal of Medical Informatics*. 2024;186:105425.

9. Venturini M, Van Keilegom I, De Corte W, Vens C. Predicting time-to-intubation after critical care admission using machine learning and cured fraction information. *Artificial Intelligence in Medicine*. 2024;150:102817.

10. Amory IA, Rashidi Khazaee P, Yousefi S. Improving sepsis mortality prediction with machine learning using full region synthetic sampling approach. *Health Science Reports*. 2025;8(11):e71556.

11. Gao J, Lu Y, Ashrafi N, et al. Prediction of sepsis mortality in ICU patients using machine learning methods. *BMC Medical Informatics and Decision Making*. 2024;24:228.

12. An S, Ye Z, Che W, et al. Development and validation of machine learning models to predict in-hospital mortality in ICU patients with sepsis and chronic kidney disease. *BMC Infectious Diseases*. 2025;25(1):1504.

13. Silveira EC, Pretti SM, Santos BA, et al. Prediction of hospital mortality in intensive care unit patients from clinical and laboratory data: a machine learning approach. *World Journal of Critical Care Medicine*. 2022;11(5):317–329.

14. Li H, Ashrafi N, Kang C, Zhao G, Chen Y, Pishgar M. A machine learning-based prediction of hospital mortality in mechanically ventilated ICU patients. *PLoS ONE*. 2024;19(9):e0309383.

15. Weaver M, Goodin DA, Miller HA, et al. Prediction of prolonged mechanical ventilation in the intensive care unit via machine learning: a COVID-19 perspective. *Scientific Reports*. 2024;14:30173.

16. Zhang G, Xie Q, Wang C, Xu J, Liu G, Su C. Intelligent alert system for predicting invasive mechanical ventilation needs via noninvasive parameters: employing an integrated machine learning method with integration of multicenter databases. *Medical & Biological Engineering & Computing*. 2024;62(11):3445–3458.

17. Hannon DM, Syed JDA, McNicholas B, Madden M, Laffey JG. The development of a C5.0 machine learning model in a limited data set to predict early mortality in patients with ARDS undergoing an initial session of prone positioning. *Intensive Care Medicine Experimental*. 2024;12(1):103.

18. Misseri G, Piattoli M, Cuttone G, Gregoretti C, Bignami EG. Artificial intelligence for mechanical ventilation: a transformative shift in critical care. *Therapeutic Advances in Pulmonary and Critical Care Medicine.* 2024;19:29768675241298918.

19. El-Rashidy N, Tarek Z, Elshewey AM, et al. Multitask multilayer-prediction model for predicting mechanical ventilation and the associated mortality rate. *Neural Computing and Applications*. 2025;37:1321–1343.