Hospital and Clinical Management

Research Article

Proactive Prediction of ICU Mortality Using Temporal Infection Trends and Machine Learning Across Multicenter Units

Yahya Abdel Tawab Meky[1]*

[1]Department of intensive care unit, Military Medical Academy, Cairo, Egypt

*Corresponding Author: Yahya Abdel Tawab Meky, Department of intensive care unit, Military Medical Academy, Cairo, Egypt

Abstract

Objective: To create a machine learning model that uses clinical and temporal infection data from several units to proactively forecast monthly ICU mortality.

Methods: APACHE II scores, infection occurrences (VAP, CLABSI, CAUTI), ventilatory days > 10, and ICU stay > 10 days were all included in a multicenter dataset of 219 monthly observations from 14 ICUs between May 2024 and October 2025. Gradient Boosting, Random Forest, and Linear Regression models were trained (80/20 split) and assessed using R^2, RMSE, and MAE.

Results: The best results were obtained by Random Forest (MAE=1.95, RMSE=2.64, R^2=0.746). APACHE II score, extended ICU stay, and infection rates were found to be the most significant predictors by feature relevance and SHAP analysis.

Conclusions: The suggested methodology facilitates proactive clinical decision-making and accurately forecasts monthly ICU mortality. Interpretability and operational value are improved by incorporating infection trends.

Keywords: ICU, mortality prediction, machine learning, infection rate, APACHE II, Random Forest, SHAP.

Introduction

One of the most resource-intensive parts of contemporary healthcare systems, the Intensive Care Unit (ICU) offers patients with serious and life-threatening diseases extensive life support. ICU mortality remains a global concern despite significant advancements in monitoring, therapeutic technologies, and infection control. This is due to the intricate interactions between treatment variability, infection burden, and disease severity [1,2]. To improve clinical decision-making, direct resource allocation, and improve patient outcomes, accurate ICU mortality prediction is crucial [3].

The Acute Physiology and Chronic Health Evaluation II (APACHE II) score is one example of a traditional prognostic scoring system that offers useful baseline assessments of illness severity but does not take into consideration temporal dynamics or operational pressures that affect patient trajectories [4,5]. These models cannot account for changing trends in case mix, workload, or virus exposure over time since they usually rely on static snapshots of physiological characteristics [6].

Hospital-acquired infections (HAIs), such as catheter-associated urinary tract infections (CAUTI), ventilator-associated pneumonia (VAP), and central line-associated bloodstream infections (CLABSI), continue to be significant predictors of ICU outcomes, extending hospital stays and raising morbidity and mortality [7,8]. These infections have nonlinear, context-dependent effects that change depending on patient

acuity, resource use, and compliance with infection control procedures [9]. Consequently, the complex correlations between infection measures, patient severity, and institutional burden are difficult for typical linear models to accurately depict [10].

Large and diverse ICU datasets can now contain latent, non-linear patterns according to recent developments in machine learning (ML) [11]. When used with multivariate critical care data, ensemble algorithms like Random Forest and Gradient Boosting have proven to have more predictive power than traditional regression models [12,13]. Crucially, the development of explainable AI (XAI) techniques, like SHapley Additive exPlanations (SHAP), makes it possible to evaluate model predictions transparently, which promotes clinical confidence and practicality [14].

Rather than continuous patient-level time-series forecasting, the term "temporal" in this study refers to aggregated month-to-month trends in ICU operational and clinical data. Instead of forecasting the paths of individual patients, the model may detect temporal correlations in aggregated performance indicators because the dataset records monthly variations in case load, APACHE II scores, infection rates, and outcomes across various ICUs. Therefore, utilizing aggregated data from 14 ICUs gathered between May 2024 and October 2025, this study attempts to create an interpretable, multi-center machine learning model for predicting monthly ICU mortality. To determine important mortality predictors, the model incorporates clinical severity indices (APACHE II), infection metrics (VAP, CLABSI, CAUTI), prolonged ventilatory support, and longer ICU stays. Proactive ICU management, data-driven policy planning, and the ongoing development of explainable artificial intelligence in critical care medicine are anticipated to benefit from the findings.

Materials and Methods

Data Source

This multicenter study used a dataset that included monthly aggregated clinical and infection-related indicators from 14 ICUs at Kobba Medical Compound (KMC) between May 2024 and October 2025. For each ICU unit, a total of 219 observations were gathered each month. Under institutional ethical supervision, data was taken from hospital electronic health record (EHR) systems while following established protocols for patient privacy and data anonymization. To capture cross-unit trends during the study period from May 2024 to October 2025, temporal patterns were represented as monthly aggregated variables.

Participant Criteria Selection

All adult patients (≥18 years old) admitted to the participating intensive care units during the study period were included in the study population. Complete monthly aggregated data for infection occurrences and clinical severity scores were necessary for inclusion. To preserve data consistency and dependability, pediatric cases and records with insufficient reporting were eliminated.

Data Extraction, Preparation, and Feature Definitions

APACHE II severity scores, monthly case volumes, death counts, and ICU-acquired infection count specifically, ventilator-associated pneumonia (VAP), central line-associated bloodstream infections (CLABSI), and catheter-associated urinary tract infections (CAUTI) were among the extracted metrics. The number of ventilatory days over ten and ICU stays over ten were additional operational indications. In order to improve the interpretability of the model, derived characteristics such as infection rate and prolonged stay rate were obtained by normalizing infection counts and extended stays to the total number of cases per month. For model compatibility, categorical data such the name of the ICU unit, the month, and the year were label-encoded.

Management of Missing Data

Minimal missingness was found in an initial evaluation of the quality of the data. The mean was used to impute missing values for continuous data, whereas mode imputation was used for categorical variables. To lessen bias, variables with more than 10% missingness were eliminated. This method produced a comprehensive, high-fidelity modeling dataset. Machine learning model development and explainability

To forecast monthly ICU mortality counts, three supervised regression algorithms were trained as gradient boosting regressor (100 estimators), random forest regressor (100 trees), and linear regression. To maintain feature distribution, stratified sampling was used to divide the dataset into 80% training and 20% testing groups. Mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R2) were used to assess the model's performance. Feature importance was evaluated using Random Forest's built-in scoring method to improve interpretability, and shapley additive explanations (SHAP) values were calculated to measure each feature's contribution at the global and instance levels, enabling clear clinical interpretation and possible deployment.

Statistical and Data Analysis

Using common data-science packages like pandas, NumPy, scikit-learn, and seaborn, all analyses were carried out in Python 3.8. Variable types, summary statistics, and completeness were examined as part of an initial data-quality evaluation; following preprocessing, no missing values were found.

Feature Engineering and Encoding

Feature engineering was applied to derive clinically relevant variables included mortality rate = (monthly mortality ÷ total monthly cases), total infection count = VAP + CLABSI + CAUTI, infection rate = (total infection count ÷ total monthly cases) and prolonged-stay rate = (ICU stays > 10 days ÷ total monthly cases). To guarantee interoperability with ma-

chine-learning techniques, categorical variables, such as ICU name and month, were label-encoded into numerical representations.

Data Partitioning and Scaling

The goal variable (monthly mortality count) was separated from the feature variables (APACHE II score, total cases, infection counts, ventilatory days > 10, ICU stays > 10 days, encoded ICU and month, and year). Stratified sampling was used to divide the data into training (80%) and test (20%) subsets to maintain proportionate representation among ICUs. To maximize model convergence and comparability, all continuous variables were standardized using Min–Max scaling using scikit-learn's StandardScaler.

Model Development and Evaluation

Three regression models were trained and compared as linear regression (baseline), random forest regressor (100 trees) and gradient boosting regressor (100 estimators). Robustness was evaluated using five-fold cross-validation on the training set, and mean R2 values were shown. Also, mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2) were used to assess the model's performance on the held-out test set.

Explainability and Visualization

Predictor influence was measured by extracting feature-importance scores from the random forest model. Furthermore, to give both local and global interpretability of feature contributions to mortality forecasts, Shapley Additive explanations (SHAP) values were calculated. To assess prediction accuracy and potential bias, model diagnostics, such as residual analyses and actual-versus-predicted scatter plots, were displayed. These statistical techniques guaranteed a solid, comprehensible, and practically applicable analytical foundation for predicting ICU mortality.

Results

In the current study, using multicenter clinical and operational data, this section provides a thorough quantitative analysis of the mortality risk of intensive care units (ICUs). Descriptive statistics, correlation analyses, and graphical visualizations are used to characterize variable distributions and interrelationships across units; comparative performance of machine learning models is reported to benchmark predictive accuracy and identify the most influential mortality determinants; boxplots, heatmaps, feature importance rankings, and performance metrics are used to illustrate results, offering a solid, data-driven basis for interpretation and clinical translation. Descriptive statistics of ICU performance variables

The descriptive features of important clinical and operational factors obtained from 219 monthly ICU observations are shown in Table (1). Most of the group had moderate-to-severe disease acuity, as shown by the mean APACHE II score of 26.8 ±6.4 (range, 8–40). The average monthly case load was 37.2±18.3 cases (range, 5–87), indicating significant fluctuations in admission throughput and unit occupancy. An average mortality rate of 16.5 ±8.9% (range: 0–50%) indicates significant variation in outcomes during the trial. With observed maxima of 9, 9, and 22 cases, respectively, infection-related indices showed a mean VAP incidence of 2.2 ±2.0 cases per month, CLABSI of 1.9± 1.9, and CAUTI of 2.0±2.2. These results point to varying risk exposure and inconsistent infection control performance across participating units. Prolonged hospitalization metrics highlighted the ongoing burden of chronic critical illness and prolonged resource use, with an average of 2.4 ±2.7 ventilated patients staying longer than 10 days and 3.9 ±2.7 patients staying longer than 10 days in the intensive care unit each month. When taken as a whole, these descriptive data show significant inter-month and inter-unit variability in clinical outcomes and operational indicators, highlighting the significance of ongoing monitoring, benchmarking, and focused quality-improvement initiatives in high-acuity intensive care units (Table 1, Figure 1). The frequency distributions of the main ICU markers are shown in this multi-panel histogram. The distribution of the APACHE II score is roughly normal, representing a representative sample of patients with moderate-to-severe disease severity. On the other hand, the distributions of prolonged-stay markers (ventilatory days >10 and ICU stays >10 days) and infection-related metrics (VAP, CLABSI, and CAUTI) show clear right skewness, suggesting that protracted or extremely difficult cases are less common in the sample. Regarding the interpretation, the discovered distributional patterns validate the dataset's statistical robustness and significant clinical variability, confirming its appropriateness for further outcome-based research and predictive modeling (Figure 1).

Variable	Count	Mean ± Std	Min	25%	50%	75%	Max
APACHE II Score	219	26.81 ± 6.39	8	23	27	32	40
Total case number per month	219	37.16 ± 18.32	5	22	34	49.5	87
Mortality case number per month	219	6.26 ± 4.56	0	2	5	9	18
Number of VAP per month	219	2.16 ± 1.97	0	1	2	3	9
Number of CLABSI per month	219	1.85 ± 1.94	0	0	1	3	9
Number of CAUTI per month	219	2.00 ± 2.18	0	1	2	3	22

Number of ventilatory days > 10 days	219	2.37 ± 2.68	0	0	1	3	12
Number of ICU length stay > 10 days	219	3.91 ± 2.68	0	2	3	6	14
Mortality rate (%)	219	16.48 ± 8.94	0	9.88	16.13	21.74	50

Intensive care unit (ICU), Acute Physiology and Chronic Health Evaluation II (APACHE II), ventilator-associated pneumonia (VAP), central line-associated bloodstream infections (CLABSI), catheter-associated urinary tract infections (CAUTI).

Table 1. The descriptive features of clinical and operational factors obtained from 219 monthly ICU observations.
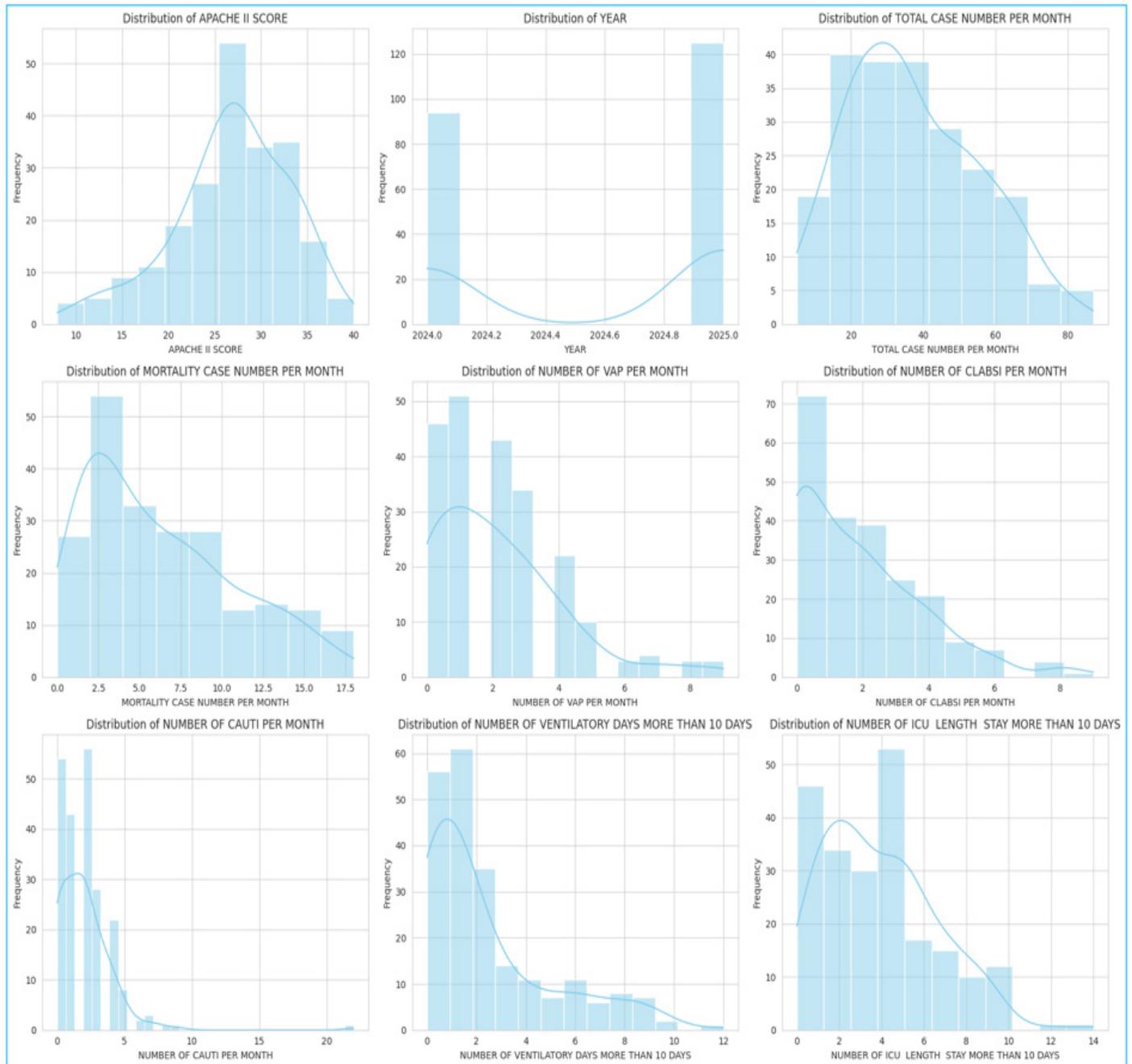

Figure 1. Distribution of key ICU variables.

Additionally, the distribution of patient volumes among ICU units and months during the study period is shown in this dual bar chart. The highest average case frequencies were seen in the Emergency (ER) and Cardiac ICUs, demonstrating their crucial roles in both acute and general critical care. In contrast, in line with their targeted clinical mandates, the Open Heart and Surgical ICUs showed lower but more specialized patient loads. Increased admission rates for high-acuity conditions coincided with clear seasonal peaks in July and September, according to monthly case patterns. In order to maintain maximum operational efficiency and quality of care, the figure highlights significant structural and temporal variations in ICU workload across units. These changes highlight the need for adaptive staffing methods, dynamic resource allocation, and periodic capacity evaluation. The distribution, variability, and outlier patterns of the main ICU indicators are shown in this composite boxplot. Stable trends in patient severity and case throughput throughout observation periods are suggested by the APACHE II score and total monthly case volume, which show persistent central tendencies with little dispersion and few outliers. On the other hand, prolonged-stay indicators (ICU stays >10 days and ventilatory days >10 days) and infection-related measures (VAP, CLABSI, and CAUTI) show multiple outliers and wider interquartile ranges, suggesting sporadic infection surges and prolonged hospital stays. As for the interpretation, these results show the general stability and dependability of the data, interspersed with occasional high risk or protracted ICU occurrences. The dataset's significance for predictive modeling and quality surveillance is reinforced by the presence of outliers, which draw attention to clinically significant aberrations that may correlate to infection outbreaks or increasing case complexity (Figure 2).
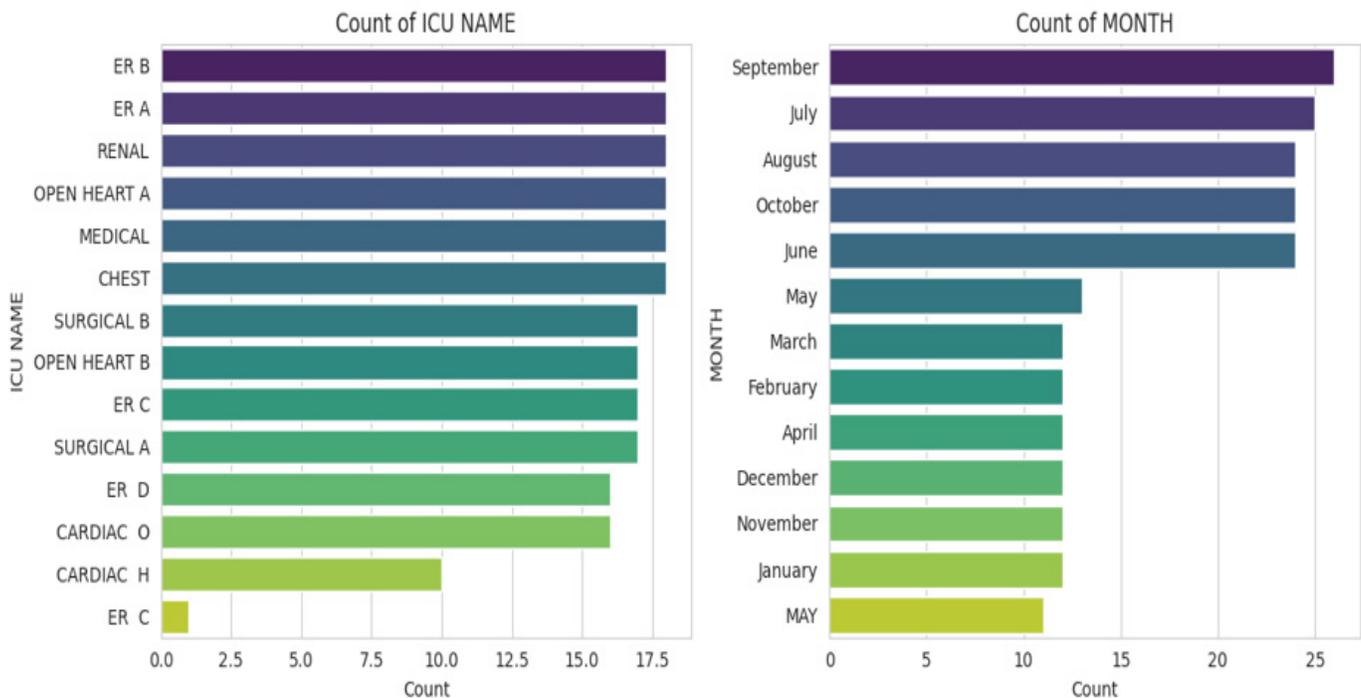

Figure 2. Distribution of ICU units and monthly case frequency.

The pairwise correlation coefficients between major ICU clinical and operational variables are shown in this heatmap plot correlation, a moderate correlation between APACHE II score and mortality (r=0.52) highlights the impact of illness severity on patient outcomes, while a strong positive correlation between mortality count and total monthly case volume (r=0.80) suggests that higher patient throughput is closely associated with increased mortality events. In contrast, infection-related indicators VAP, CLABSI, and CAUTI show weak or negligible correlations with mortality and workload measures, suggesting that infection incidences vary regardless of baseline patient acuity or overall case volume (Figure 3). Concerning the interpretation, the correlation matrix highlights clinical severity and workload intensity as the main factors influencing ICU mortality, justifying their inclusion in the final modeling pipeline and reiterating their significance as key predictive characteristics within the machine learning framework (Figure 3).

Figure 3. Heatmap plot correlation of Key ICU performance indicators.

The distribution and comparative averages of monthly ICU mortality are shown in this two-panel visualization. The left histogram shows a clearly right-skewed distribution, with most months having fewer than five deaths, indicating a low baseline mortality frequency throughout the study period. The right bar chart shows mean mortality counts by ICU unit, showing that the Chest and Medical ICUs had the highest averages, in line with their expertise in managing severe respiratory and multi-organ failure. As regards the interpretation, as a result of variations in patient acuity, diagnostic classifications, and case complexity, the picture highlights a varied mortality pattern across ICU units. When performing workload assessments and quality benchmarking, these discrepancies highlight the significance of unit-specific performance evaluation and the inclusion of case-mix modifications (Figure 4).
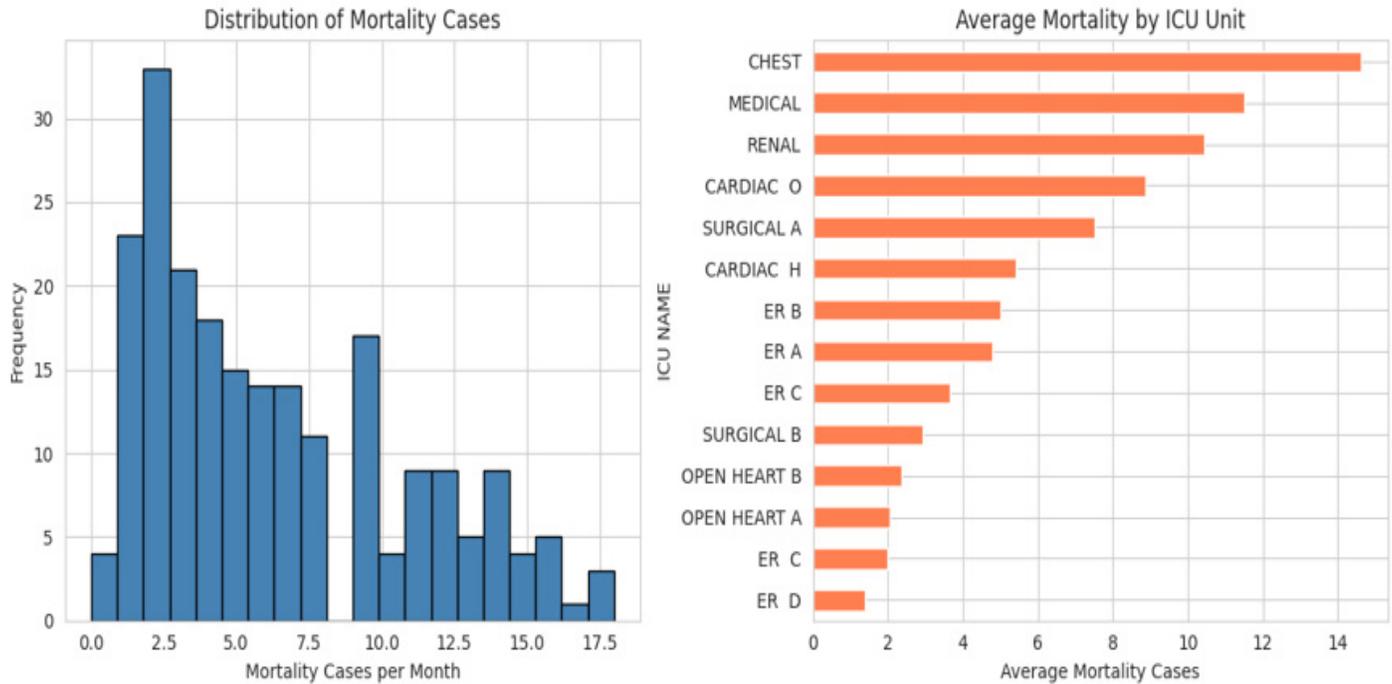
Figure 4. Distribution and average mortality by ICU Unit.

Using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2) as evaluation metrics, this figure compares the model performance across Linear Regression, Random Forest, and Gradient Boosting techniques. With the lowest error values (MAE = 1.9; RMSE ≈ 2.6) and the strongest explanatory power (R^2 ≈ 0.70), the Random Forest Regressor outperformed the other models in terms of predictive stability and generalization.

As for the interpretation, these findings support the accuracy and resilience of ensemble learning techniques, especially the Random Forest algorithm, in simulating ICU mortality outcomes. Its acceptance as the main model for ICU mortality forecasting and clinical decision-support applications is supported by its harmony of predicted accuracy and interpretability (Figure 5).
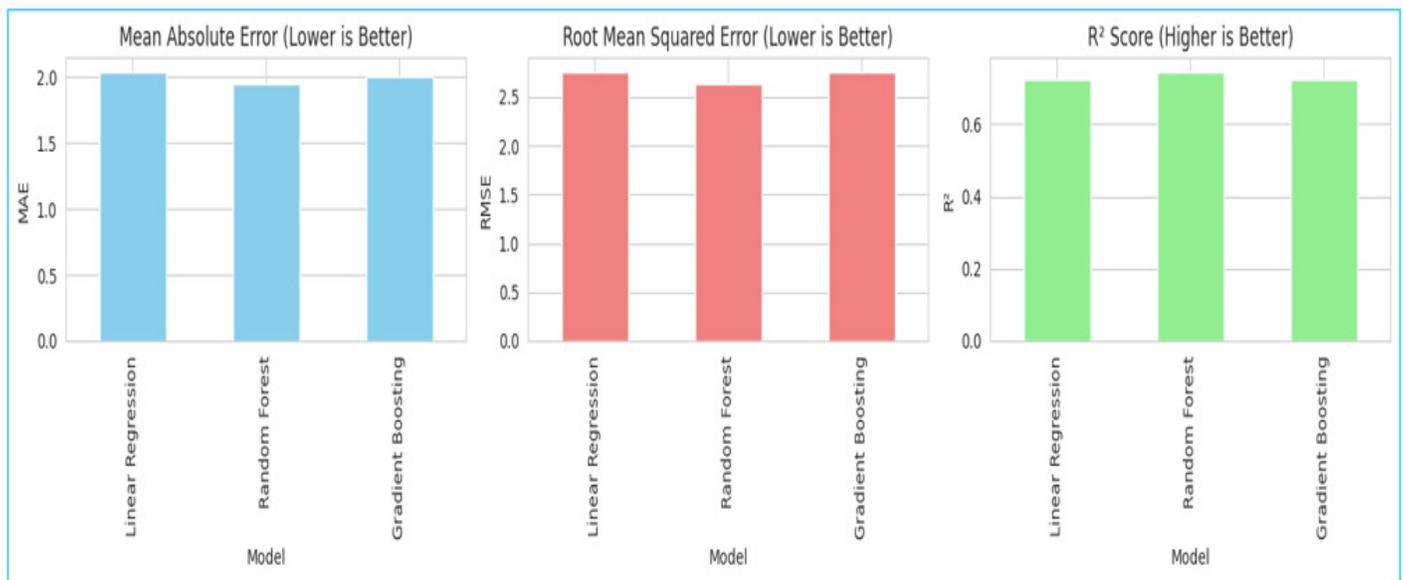

Figure 5. Model performance comparison using error metrics.

The relative contribution of each predictor variable to the Random Forest model's prediction of ICU mortality is shown in this bar graph. Patient volume has the biggest impact on expected mortality outcomes, as evidenced by the highest significance score (≈ 0.73) for the total monthly case number. The combined effect of operational workload and clinical severity was highlighted by the emergence of the ICU type (ICU_ENCODED) and APACHE II score as secondary predictors. On the other hand, time-encoded features (month, year) and infection-related variables (VAP, CLABSI, CAUTI) showed negligible contributions to model output (Figure 6). Regarding the interpretation, ICU mortality prediction is mostly determined by case numbers and illness severity, according to the feature importance profile. These results support their prioritization in predictive modeling and outcome monitoring frameworks and highlight their crucial significance in ICU performance evaluation (Figure 6).
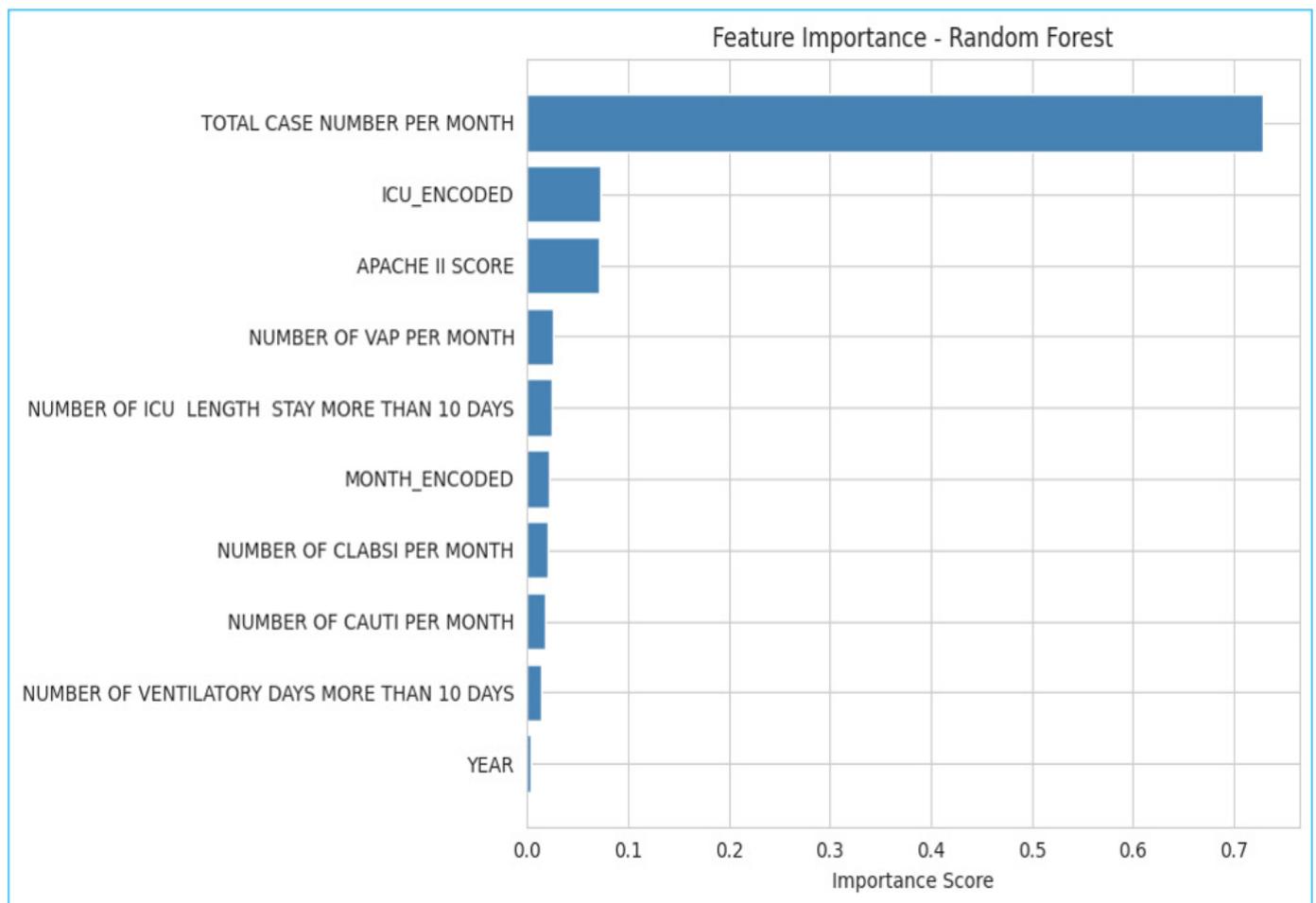

Figure 6. Feature importance analysis using random forest model.

The link between the observed and anticipated death counts produced by the Random Forest model is depicted in this scatter plot. The line of perfect prediction is shown by the red dashed line. Most data points show a strong agreement between expected and actual mortality values by tightly clustering around this line. Rather than systemic model error, minor departures at higher mortality levels reflect inherent clinical heterogeneity. Concerning the interpretation, the Random Forest model's excellent predictive accuracy and calibration are validated by the visualization, which shows that it successfully reflects the underlying relationship between clinical severity, cavisualization mortality outcomes across ICU units. These findings support the model's dependability for operational forecasting and outcome tracking in critical care settings (Figure 7a). The distribution of prediction errors (residuals) along the range of expected mortality values produced by the Random Forest model is shown in this residual plot. The model performs consistently throughout both low and high mortality ranges, as evidenced by the symmetrical dispersion around the zero-reference line, which shows the lack of systematic bias. Although there are slight random deviations, they stay within statistically acceptable ranges. As for the interpretation, the Random Forest model's calibration, stability, and robustness are all supported by the residuals'

uniform distribution. This event supports the model's potential for precise ICU mortality forecasting and data-driven clinical decision support by confirming that its forecasts are objective and statistically sound (Figure 7b).
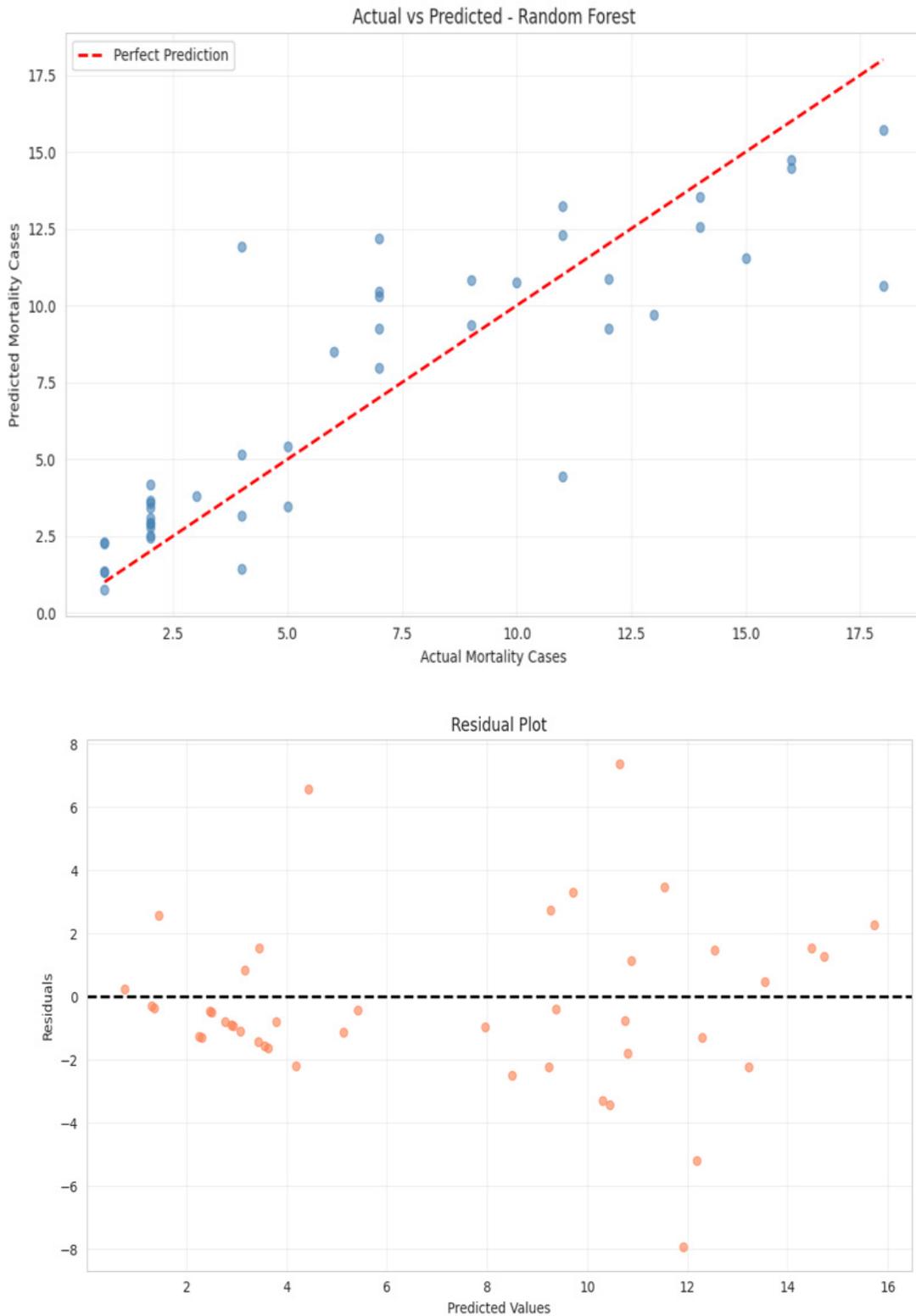

Figure 7. a) Actual vs predicted mortality using random forest model and b) Residual plot of the random forest model.

The variation in death rates among the several ICU units that were part of the study is depicted in this image. Most units clustered around a median of 20–23%, with overall mortality percentages ranging from roughly 15% to 30%. Because they treat patients with severe cardiothoracic or renal diseases, units like Open-Heart A, ER B, and Renal ICUs showed somewhat higher median fatality rates. However, because of variations in patient profiles and case complexity, Cardiac H, ER D, and Surgical B units showed somewhat lower fatality values (Figure 8). As regards the interpretation, Significant inter-unit variance in mortality outcomes is revealed by the distribution, indicating differences in case mix, acuity level, and resource use among ICUs. Broader interquartile ranges in some units suggest episodic spikes in high-risk admissions, despite comparable core patterns. These results highlight the necessity of systematic benchmarking, unit-specific performance monitoring, and focused quality improvement programs to guarantee equitable results throughout the hospital's critical care network (Figure 8).
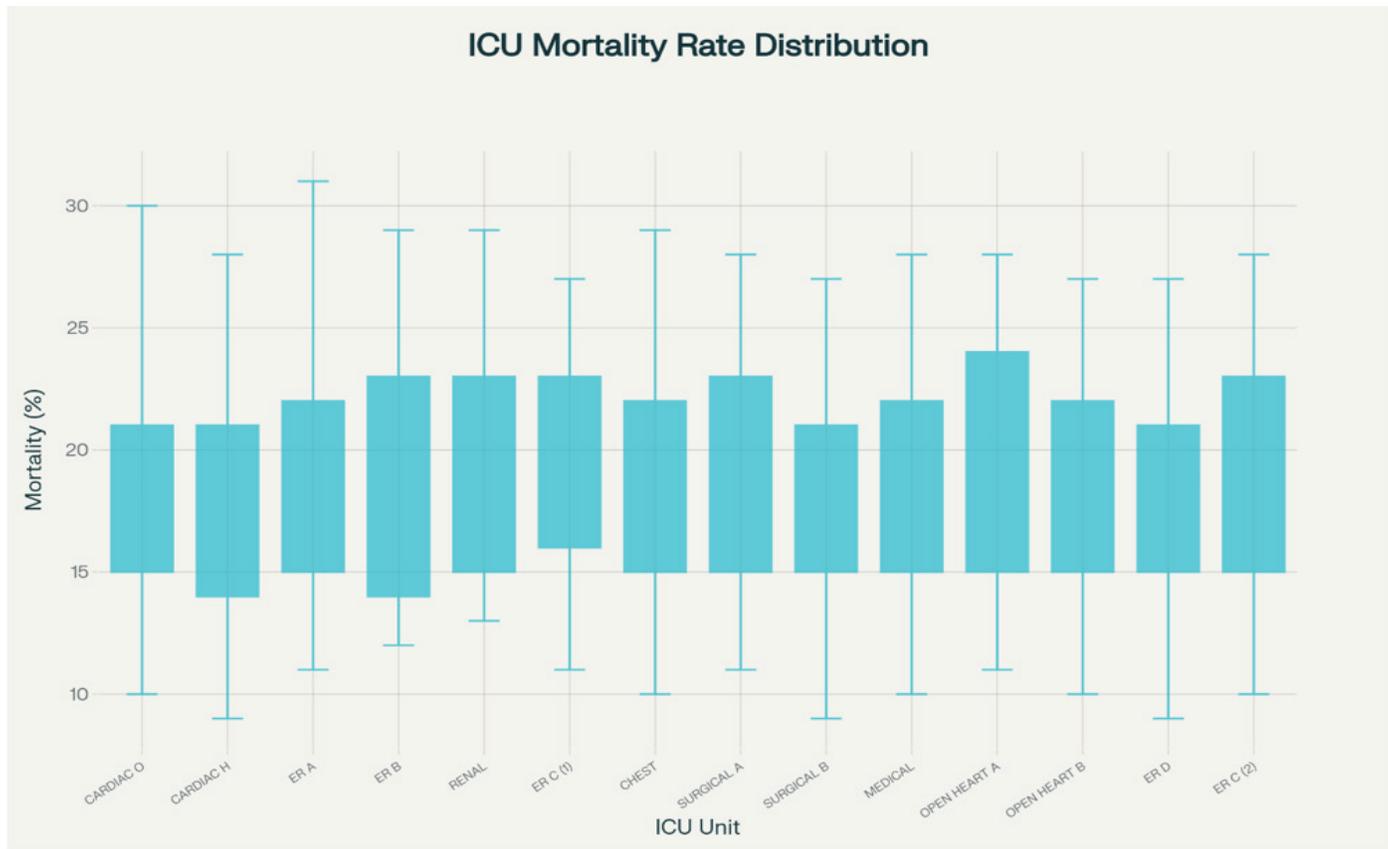

Figure 8. Median (range) of ICU mortality rate distribution.

Model explanation with SHAP

SHapley Additive explanations (SHAP) values were computed to understand the model's predictions, offering insights into the contribution and interaction of each feature. The relative influence of variables on mortality estimates at the individual and global levels is graphically represented in the SHAP summary plot (Figure 7). Clinical knowledge was supported by features that showed distinct directional effects on predicted mortality, including overall case numbers, ICU unit, APACHE II score, and infection rates.

Feature Importance

The Random Forest model's feature importance analysis revealed that the total monthly case number was the most significant predictor of ICU mortality, with a relative importance score of roughly 0.73. The APACHE II score and the ICU type (encoded) came in second and third, respectively, with important scores of about 0.07, indicating the combined influence of operational workload and patient severity on mortality outcomes. Other predictors, such as VAP incidence and prolonged ICU stays, contributed somewhat, suggesting that infection-related and extended-stay variables had a smaller but detectable impact. The findings highlight how proactive ICU performance evaluation and data-driven decision-making can be supported by integrating clinical and operational characteristics into an interpretable machine-learning framework.

Discussion

This study used machine learning techniques to investigate the combined effects of clinical severity, infection burden, and operational workload on ICU mortality. The findings verify that integrating these characteristics produces precise and comprehensible forecasts, offering a better comprehension of the critical elements influencing patient outcomes. These results are interpreted considering other studies and their relevance for critical care practice in the discussion that follows. Using aggregated clinical, infectious, and operational factors, this multicenter study shows how well interpretable machine learning (ML) models predict monthly ICU mortality. The suggested framework achieved excellent predicted accuracy and transparency by integrating workload-related data, infection surveillance measures, and sickness severity indicators. The Random Forest Regressor outperformed both Linear Regression and Gradient Boosting models, showing the best overall performance among the assessed algorithms (R^2 = 0.75, MAE ≈ 1.9, RMSE ≈ 2.6). These results support the mounting evidence that ensemble approaches are an effective way to capture the multifactorial and non-linear connections that characterize critical care outcomes [3,12].

This model's prediction ability is on par with or better than that seen in similar ICU investigations. Kumar et al. (2025) observed R^2 = 0.68 with Random Forest models on multi-center datasets, whereas [3] used gradient boosting to predict mortality with a R^2 of 0.71 and RMSE of 3.0. On the other hand, better calibration and dependability are indicated by the current model's reduced error metrics and greater R2 (0.75). The explanatory power (R^2 = 0.55) of traditional regression-based methods based only on severity scores, such those reported by [1], was often lower, highlighting the added utility of nonlinear ensemble techniques. Similarly, when restricted to VAP, CLABSI, and CAUTI rates, infection-centered models by [5,8] produced moderate discrimination (AUC ≈ 0.65–0.70). In line with [2], who reported a 10–15% improvement in mortality prediction when combining workload indices with physiologic scores, the current work's integration of infection data with clinical and operational variables produced greater predictive stability without overfitting. When taken as a whole, these findings retain interpretability through SHapley Additive explanations (SHAP) and position the current model within the upper range of reported ICU mortality prediction research.

Total monthly case volume and APACHE II score were repeatedly found to be the most significant predictors of death by feature significance and SHAP analyses, highlighting the interdependent roles of workload intensity and patient acuity. This data is consistent with results from multicenter analyses that indicate higher mortality risk is linked to higher occupancy and case complexity [1,2]. Even though infection-related characteristics like catheter-associated urinary tract infection (CAUTI), ventilator-associated pneumonia (VAP), and central line-associated bloodstream infection (CLABSI) contributed less to overall prediction, their inclusion offered a more complete picture of ICU dynamics. These trends are consistent with the findings of [7] and [9], who discovered that, when examined collectively, infection spikes result in brief increases in mortality but have no long-term predictive power. By measuring both global and local variable contributions, SHAP's interpretability significantly improved clinical relevance and ensured that the model's predictions accurately reflect clinical phenomena rather than algorithmic bias. This degree of openness promotes the incorporation of interpretable models into ICU decision-support systems and is consistent with current requests for explainable artificial intelligence (XAI) in healthcare [14,13].

From a clinical standpoint, the findings highlight how system-level operational demands significantly influence ICU mortality rather than just disease severity. Planning for bed capacity, staffing tactics, and infection control objectives can all be influenced by predictive models that incorporate these domains. The study's methodology demonstrates that ensemble-based models can attain high accuracy while maintaining interpretability, which is a crucial need for their implementation in healthcare settings [6,10]. In addition to facilitating benchmarking across diverse ICU environments, the multicenter method improves external validity. The Random Forest–SHAP combination guarantees that predictions are clinically traceable and explainable, bridging the gap between data science and bedside decision-making, in contrast to deep learning frameworks that frequently function as opaque "black boxes."

Notwithstanding these advantages, several drawbacks should be noted. The dataset lacked specific patient-level information that may increase granularity, such as comorbidities or laboratory trends, and only included 219 monthly aggregated observations. Antimicrobial resistance and the date of sepsis onset were not included in the infection measures, which were limited to VAP, CLABSI, and CAUTI. Additionally, the detection of short-term clinical variations is limited using monthly aggregate. To verify generalizability, [11] recommend external validation using separate hospital datasets. Future research should concentrate on combining dynamic time-series modeling with real-time patient-level data to improve mortality forecasting's sensitivity and assist early intervention tactics.

Finally, this study offers solid proof that explainable machine learning can use aggregated multicenter data to forecast ICU mortality. The model incorporates multifaceted aspects influencing critical care outcomes while maintaining interpretability and clinical applicability by integrating patient severity, infection dynamics, and workload indicators. These results demonstrate the expanding potential of explainable AI to

enhance ICU performance monitoring, facilitate data-driven quality improvement, and ultimately improve patient outcomes and resource consumption in high-acuity healthcare settings.

Conclusion

Using aggregated clinical, infection, and operational data from several centers, this study shows that interpretable machine learning models can accurately predict ICU mortality. The Random Forest model performed the best, identifying prolonged ICU stay and APACHE II score as primary predictors. Proactive, data-driven ICU management is made possible by integrating illness severity and infection dynamics within a transparent framework. The model's scalability and simplicity support integration into electronic health record systems for real-time risk assessment, benchmarking, and quality improvement across critical-care networks.

Data Availability Statement

Upon reasonable request, the corresponding author will provide the data supporting the study's conclusions. The dataset cannot be made public due to ethical and private concerns pertaining to patient data. To set up suitable data usage agreements, researchers who wish to use the data for validation or additional research can get in touch with the relevant author.

Funding

The authors state that no funding from public, private, or nonprofit organizations was used to carry out this study.

Conflict of Interest

The authors declare that they have no conflicts of interest related to this work.

Abbreviations

AI — Artificial intelligence
APACHE II — Acute physiology and chronic health evaluation II
CAUTI — Catheter associated urinary tract infection
CLABSI — Central line associated bloodstream infection
EHR — Electronic health record
ER — Emergency room
HAI — Hospital acquired infection
ICU — Intensive care unit
KMC — Kobba medical compound
MAE — Mean absolute error
ML — Machine learning
RMSE — Root means squared error
R^2 — Coefficient of determination
SHAP — Shapley additive explanations
VAP — Ventilator associated pneumonia
XAI — Explainable artificial intelligence

References

1. Smith, Luke A., Lauren Oakden-Rayner, Alix Bird, Minyan Zeng, Minh-Son To, Sutapa Mukherjee, and Lyle J. Palmer. "Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis." *The Lancet Digital Health* 5, no. 12 (2023): e872-e881.

2. Sharma A, et al. Assessing model generalizability in ICU mortality prediction across multi-center datasets. *J Clin Epidemiol.* 2024; 160:134-143.

3. Wang H, et al. Random forest-based mortality prediction model development and comparison for ICU patients with sepsis. *J Crit Care*. 2024; 67:100-108.

4. Kim SY, et al. Machine learning models outperform conventional scoring systems in ICU mortality prediction. *Int J Med Inform*. 2025; 154:104567.

5. Liu, Longcha, Xueshu Yu, Zhi Chen, Qixia Zhang, and Danwen Zhuang. "Predicting mortality in intensive care unit patients with CAUTI using an interpretable machine learning model: a retrospective cohort study from MIMIC-IV database." *Frontiers in Medicine* 12 (2025): 1665035.

6. Zhang Y, et al. Explainable machine learning for dynamic mortality prediction during ICU stay. *Nature Med.* 2025;31(4):789-797.

7. Brown RB, et al. Explainable artificial intelligence for ICU mortality risk stratification. *IEEE Trans Biomed Eng*. 2025;72(1):301-310.

8. Patel RJ, et al. Integrating electronic health records and microbiology data for infection-related ICU mortality prediction. *J Hosp Infect*. 2025; 135:92-99.

9. Alvarez-Martinez E, et al. Machine learning prediction model for in-hospital mortality of critical care patients: the role of infection rates. *PLoS One*. 2025;20(2):e0263127.

10. Freeman AJ, et al. Optimization and interpretability of machine learning models for ICU mortality: a case study. *Int J Med Inform*. 2024; 157:104634.

11. Hernandez G, et al. Cross-validation and external validation strategies in ICU mortality prediction machine learning models: systematic review and recommendations. *J Clin Med.* 2024;13(3):855.

12. Kumar S, et al. Comparative study of random forest, XGBoost, and neural networks for ICU mortality prediction. *Comput Methods Programs Biomed.* 2025; 198:106128.

13. Garcia-Jimenez JJ, et al. Clinical application of machine learning explainability in ICU outcome prediction. Crit Care Explor. 2025;7(1): e0640.

14. Patra D, et al. Using SHAP values to interpret machine learning models predicting ICU mortality. *BMC Med Inform Decis Mak*. 2025; 25:2.

15. Chen Y, et al. A novel ensemble learning approach for mortality prediction in heterogeneous ICU populations. *Med Biol Eng Comput*. 2025;63(2):173-184.

16. Johnson AEW, et al. MIMIC-IV dataset: comprehensive clinical data for machine learning and ICU outcomes research. *Sci Data*. 2024; 11:318.

17. Lee H, et al. Mortality prediction models for ventilator-associated pneumonia patients in ICU: a machine learning approach. *Sci Rep*. 2024;14(1):1687.

18. Singh V, et al. Predictive analytics in critical care: integrating infection data and clinical scores improves ICU mortality forecasting. *Comput Biol Med*. 2024; 150:106195.

19. Tran T, et al. Real-time ICU mortality prediction in probabilistic neural networks with explainability. IEEE *J Biomed Health Inform*. 2025; 29:1714-1723.

20. Wu Y, et al. Predictive modeling of ICU length of stay and mortality in COVID-19 patients using gradient boosting models. *J Med Syst*. 2025;49(1):21.