BPR MCS | Bulletin of Pioneering Researches of Medical and Clinical Science

Bulletin of Pioneering Researches of Medical and Clinical Science

Available online: https://bprmcs.com
2026 | Volume 6 | Issue 1 | Page: 215-227

Mixed-Effects Random Forest Modeling for Dynamic ICU Mortality Prediction Using Repeated Clinical Measurements

James Walker[1]*, Olivia Harris[1]

[1]Department of Clinical and Medical Sciences, University of Edinburgh, Edinburgh, United Kingdom.

Abstract

Researchers created a machine learning-based system for predicting mortality risk to overcome difficulties in choosing suitable participants for clinical trials conducted in the Intensive Care Unit (ICU). This algorithm combines Red blood cell Distribution Width (RDW) information with various demographic factors to forecast ICU mortality, working together with established ICU mortality scoring tools such as the Simplified Acute Physiology Score (SAPS). The present study introduces a machine learning-based prognostic scoring system for mortality that integrates RDW with other readily obtainable patient variables in the ICU. The new algorithm, called Mixed-effects logistic Random Forest for binary data (MixRFb), combines Random Forest (RF) classification with a mixed-effects model tailored for binary outcomes while properly handling repeated measurements. Comparisons of performance were made between standard RF and the new MixRFb approaches that used only SAPS scoring, supplemented by a descriptive receiver operating characteristic curve assessing RDW's capacity to predict mortality. When RDW and other covariates were included, the MixRFb model performed better than the version based solely on SAPS, delivering an area under the curve of 0.882 instead of 0.814. Variable importance plot analysis revealed that age and RDW were the strongest predictors of ICU mortality. The MixRFb algorithm is more effective at forecasting in-hospital mortality and identifies age and RDW as the key predictive factors. Using this tool could streamline participant selection for clinical trials, leading to stronger trial results and better adherence to ethical principles. Subsequent studies should prioritize making the algorithm more robust, extending its applicability across a wider range of clinical environments and patient populations, and adding additional predictive variables to enhance the precision of patient selection.

Keywords: Mortality prediction, Machine learning, Intensive care unit, Clinical trial, Patient recruitment

Corresponding author: James Walker
E-mail: james.walker@gmail.com

Received: 01 February 2026
Revised: 22 April 2026
Accepted: 29 April 2026

How to Cite This Article: Walker J, Harris O. Mixed-Effects Random Forest Modeling for Dynamic ICU Mortality Prediction Using Repeated Clinical Measurements. Bull Pioneer Res Med Clin Sci. 2026;6(1):215-27. https://doi.org/10.51847/Sc1T0NsLNl

Introduction

The Intensive Care Unit (ICU) constitutes a highly intricate environment dedicated to treating patients facing critical, often life-endangering illnesses. Carrying out Randomized Controlled Trials (RCTs) in such a setting presents notable difficulties stemming from wide differences in patient status, the pressing demand for immediate actions, and the serious ethical issues associated with research involving severely ill individuals [1]. A crucial factor in the success of ICU-based RCTs is accurately identifying patients who stand to gain the most from the tested intervention [2]. These trials frequently experience elevated mortality, which in turn leads to higher dropout rates and weaker reliability of the findings. Patients involved commonly deal with serious conditions, including sepsis, Acute Respiratory Distress Syndrome (ARDS), or Ventilator-Associated Pneumonia (VAP).

Although strict inclusion and exclusion criteria are applied, imprecise diagnoses often result in heterogeneous and inconsistent groups of study participants in ICU RCTs [3]. François *et al*. [3] review the repeated shortcomings observed in ICU trials and underline the importance of developing improved methods for stratifying patients to achieve better trial success. In a different investigation, Ali *et al*. [4] draw attention to the challenges of ensuring uninterrupted care within ICU environments and how this affects both patient recovery and the overall reliability of trials. Medical professionals running ICU trials must navigate numerous hurdles as they seek to promote scientific progress while protecting vulnerable participants from unnecessary risks [4]. Additional complications arise from ethical dilemmas, such as decisions to withdraw treatment and the process of obtaining informed consent from patients in critical condition [5]. This multifaceted situation clearly demonstrates the need for a reliable, practical mortality risk assessment tool to support appropriate participant selection in clinical trials [6].

Beyond that, a precisely calibrated mortality risk-scoring system designed specifically for ICU use could markedly enhance clinical trial planning by helping select individuals with the highest potential to respond to new treatments. For example, in studies focused on sepsis, such a scoring tool could identify patients more likely to live long enough to derive meaningful benefits from an innovative treatment, ultimately boosting both trial effectiveness and the quality of participant selection [7]. Similarly, for ARDS-related trials, a flexible mortality risk score enables the detection of patients at pivotal stages where specific interventions could deliver maximum benefit, supporting prompt and focused care that may lead to improved recovery [8]. Such a risk assessment tool could likewise play a valuable role in trials addressing Acute Kidney Injury (AKI) or ventilator-associated pneumonia by making certain that resources are directed toward those who would benefit most, thereby promoting more efficient use of resources and upholding higher ethical standards during the trial [9].

Established ICU mortality prediction tools, including the Acute Physiology and Chronic Health Evaluation (APACHE) [10], the Simplified Acute Physiology Score (SAPS) [11], and the Mortality Probability Models (MPMs) [12], have been widely applied to estimate patient prognosis and inform clinical decisions. However, these traditional systems come with notable drawbacks, especially when dealing with the diverse and constantly shifting patient groups typical in ICU settings [13]. For instance, despite its popularity, the APACHE score often fails to adequately reflect changes in a patient's condition over time, thereby reducing the accuracy of mortality forecasts [14]. Additionally, when screening individuals for participation in clinical trials, clinicians would benefit greatly from a fast, user-friendly instrument, as the extensive time and effort required for comprehensive evaluations can pose serious challenges in the fast-paced environment of an ICU [3]. Conventional methods such as SAPS [11] further depend on an extensive array of inputs, including arterial blood gas results and other detailed physiological data that may not be instantly accessible. This level of complexity can slow the identification of eligible candidates and may delay trial enrollment, particularly in facilities facing resource constraints.

In recent years, interest has grown in combining dynamic variables with Machine Learning (ML) techniques to enhance mortality prediction within the ICU [15]. Compared with conventional statistical models, ML approaches are better equipped to manage the complex, high-dimensional data commonly found in ICU settings. They generate more precise, time-sensitive predictions by detecting nonlinear relationships and interactions that standard models often miss [16].

Multiple published studies demonstrate the potential of sophisticated ML algorithms in this domain. For example, Makino *et al*. [17] constructed an artificial intelligence system that predicts the advancement of diabetic kidney disease by leveraging large-scale data, underscoring ML's ability to improve clinical decision-making. Likewise, Li *et al*. employed a machine learning approach to forecast in-hospital mortality among ICU patients with heart failure, reinforcing the usefulness of ML for sharpening mortality risk assessment [18].

In this broader context, adding Red blood cell Distribution Width (RDW) as a predictive variable within ML models may offer a meaningful step forward in ICU mortality prediction [19]. RDW quantifies red blood cell volume heterogeneity and is a standard parameter obtained from routine complete blood count analysis. Its value as a predictor of mortality and serious clinical events has been investigated in several studies [20-22]. In addition, RDW is a low-cost indicator of anisocytosis that can be easily assessed through inexpensive blood tests. It mirrors various acute and chronic conditions that influence the likelihood of death in ICU patients [23, 24]. Although incorporating RDW could boost the accuracy of ICU mortality forecasts, many conventional tools described in the literature that utilize this marker [25] do not adequately exploit the dynamic data collected throughout a patient's ICU admission [24, 26].

The present study introduces a machine learning-based prognostic scoring system for mortality that integrates RDW with other readily obtainable patient variables in the ICU. By using repeated measurements, the system creates a practical and user-friendly tool for risk profiling that can be easily applied in daily ICU practice. The proposed algorithm, Mixed-effects logistic Random Forest for binary data (MixRFb), combines the advantages of random forests with mixed-effects modeling to address the challenges posed by repeated observations in ICU data.

While traditional random forest models perform well at predicting clinical outcomes in the ICU [27], they assume independence among observations, which can lead to biased results in hierarchically structured data [28]. The mixed-effects element of the MixRFb algorithm explicitly accounts for correlations within individual patients by including random effects, thereby capturing the variability inherent in repeated measures such as RDW [29].

Nevertheless, mixed-effects models, on their own, have often shown limited predictive performance on longitudinal data compared with machine learning techniques [29, 30]. The MixRFb approach developed here seeks to address fixed effects (those consistent across patients) simultaneously and random effects (those specific to each individual) while capitalizing on the superior predictive capacity of random forest modeling [28]. This strategy is especially appropriate for ICU environments, where longitudinal clinical information plays a central role in mortality forecasting [31]. The resulting tool is designed to aid ICU patient selection and to manage repeated-measurement data collected during an ICU stay.

Materials and Methods

Data

The dataset employed to train the model comprises 286 patients who remained in the ICU for at least 48 h. These patients were admitted to the ICU of the University Hospital of Ferrara between August 2016 and December 2017. Clinical variables were recorded daily during the first five days of ICU admission, underscoring the longitudinal nature of the data.

Patients expected to remain in the ICU for at least 48 h were considered eligible. Exclusion criteria included age below 18 years, any history of hematological disorders, and pregnancy. Clinical and demographic information was collected daily. Written informed consent was secured from all patients able to provide it or from their legal representatives [23].

Data collection received approval from the local Ethics Committee (CE AVEC), as detailed by Fogagnolo *et al*. [23], under protocol number 160699, approved on 14 July 2016.

Study size

Initial statistical evaluations suggested that enrolling approximately 280 patients would provide 80% power to detect at least a 10% improvement in mortality prediction accuracy between the new MixRFb model and established scores such as SAPS, at a 5% significance level.

Descriptive statistics

Categorical variables are summarized by absolute and relative frequencies stratified by ICU mortality status. Quantitative variables are presented as median values with interquartile ranges. For each variable, the Odds Ratio (OR), 95% Confidence Intervals (CIs), and p-value derived from univariable logistic regression are also reported.

Machine learning models

The Mixed Effects Random Forest for binary data (MixRFb) was selected as the core method for building the ML-based tool [32]. This algorithm integrates Random Forest (RF) with mixed-effects modeling to handle repeated-measures data appropriately. It begins by applying an RF-based procedure that constructs multiple decision trees to capture intricate relationships within the data. Subsequently, a mixed-effects model is incorporated to account for the correlation structure of repeated observations, while simultaneously adjusting for fixed effects (parameters shared across all observations) and random effects (parameters that differ between individuals). Although classical mixed-effects models can accommodate hierarchical data structures, they depend on predefined linear associations and specified interaction terms, which restricts their ability to model complex patterns. In comparison, the MixRFb framework leverages the random forest structure to detect nonlinear relationships and higher-order interactions automatically [33].

Multiple models were constructed and evaluated through internal validation procedures. In the course of this work, missing values were addressed by applying the Multivariate Imputation by Chained Equations (MICE) approach [34] (refer to Section 2.5.2 for additional details).

(1)

Model A: MixRFb model with RDW serving as one of the main predictors, supplemented by the following input variables: (a) age, (b) gender, (c) time (measured in days of ICU stay), and (d) presence of any comorbidity. Covariates for the MixRFb algorithm were selected based on their established clinical significance and on their frequency of recording in everyday ICU care. Age is a well-known factor linked to ICU mortality because it reflects overall patient frailty and physiological imbalance at admission [35]. RDW was selected since it signals widespread inflammation and oxidative damage, both of which strongly influence prognosis in critically ill individuals [26]. Comorbidities were factored in to adjust for patients' pre-existing medical conditions that can alter survival chances [36]. Specifically, this analysis considered the existence of at least one of the following: diabetes [37], cardiovascular disease [38], or respiratory disease [39]. These were chosen because they have clear associations with death in the ICU, remain easy to interpret clinically, and reduce the burden of extensive data gathering in fast-paced critical care settings. Gender was incorporated to check for any outcome differences

between males and females, while ICU length of stay helped capture how a patient's condition evolves [40].

(2)

Model B: MixRFb model that relies on SAPS as a central predictor. SAPS is widely recognized as a reliable tool for forecasting ICU mortality, as it effectively summarizes a patient's initial severity and key physiological disruptions [41].

(3)

Model C: Conventional Random Forest model that includes RDW as a predictor along with the same additional features: (a) age, (b) gender, (c) time (days of ICU stay), and (d) any comorbidity.

Model training validation workflow

Figure 1 presents a flowchart outlining the complete process for training and internally validating the machine learning models via bootstrap resampling. The workflow starts from the full original dataset, which is then repeatedly resampled via bootstrapping and partitioned into a training portion (60%) and a testing portion (40%). Within each bootstrap cycle, missing data are first imputed in the training portion, then the model is fitted on the training data and subsequently assessed on the held-out test data. This entire sequence is executed across 1000 independent bootstrap replications. Upon completing all replications, the performance statistics from each run are pooled to generate the final results. The diagram employs different colors to mark distinct phases: gray indicates the starting setup, blue represents the repeated operations, and red denotes the concluding aggregation and summarization steps.

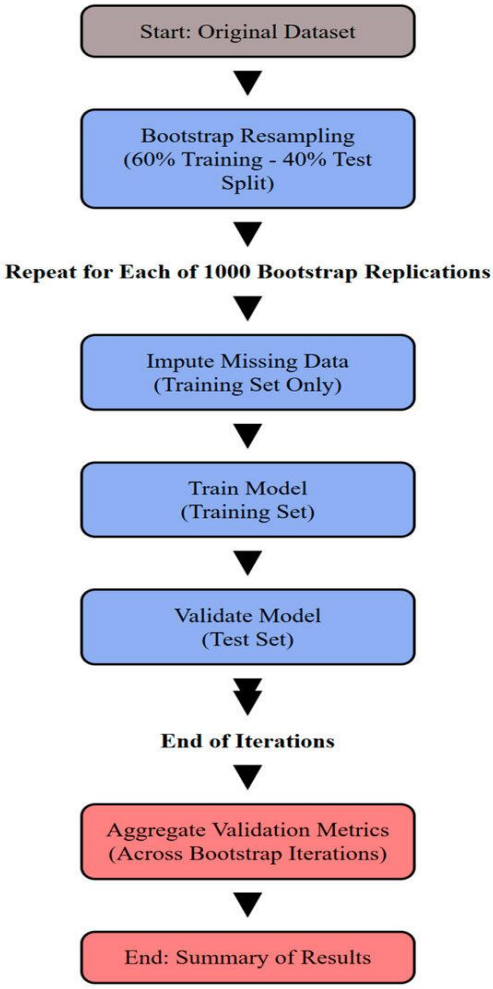


Figure 1. Model training and validation flowchart.

Each stage is described in greater detail in the sections that follow.

Model validation via bootstrap resampling

All developed models were subjected to internal validation comprising 1000 separate bootstrap resampling runs to compare their predictive abilities reliably.

During each of the 1000 bootstrap cycles, the model was fitted on a bootstrap-derived training subset containing 60% of the observations and then tested on the remaining Out-Of-Bag (OOB) cases that had not been included in that sample **(Figure 1)**. This repeated resampling strategy allows every data point to contribute to both model development and evaluation across runs, yielding robust performance estimates without altering the original proportions of survival and mortality cases [42].

Handling missing data

Missing data were managed using the MICE imputation method [34]. This technique repeatedly estimates missing values while respecting the statistical relationships among all variables and is suitable for both numeric and categorical inputs. To counterbalance the imbalance caused by the lower number of non-survivors, the minority

class was artificially increased via oversampling during model training only.

Importantly, the MICE procedure was applied only to the training subset within each bootstrap iteration, after the split had been made. This timing prevents any unintentional leakage of information from the test data. Once imputation was complete, the model was trained on the now-complete training set, and its predictions were evaluated on the untouched test set **(Figure 1)**.

Measures of performance

After completing all 1000 bootstrap iterations, the performance indicators obtained across replications were combined into summary statistics. In addition to the Area Under the ROC Curve (AUC), the F1-score was also calculated. The F1-score, defined as the harmonic mean of precision and recall, provides a more balanced view of model quality, especially valuable when dealing with uneven class distributions, where simple accuracy figures can be deceptive. Both (a) training-phase and (b) bootstrap-validated performance results—specifically the AUC and F1 values—are reported for each model. Finally, the best-performing model is benchmarked against a straightforward descriptive ROC analysis that uses only RDW and SAPS as standalone mortality predictors.

Variable importance

The Variable Importance Plot (VIP) was generated to show the average reduction in prediction accuracy that results when any given predictor is excluded from the random forest part of the model.

Predictor influence was examined more extensively with a multi-way importance diagram. This chart merges the mean drop in the Gini impurity measure — which reflects how much each feature improves the model's overall accuracy — with how often that feature is chosen for the initial splits at the root nodes of the decision trees, thereby underscoring its impact on the most critical early decisions. In addition, statistical significance was assessed for each variable through a one-sided binomial test grounded in the binomial distribution Bin (total nodes, P(node splits on Xj)), where the probability assumes uniform random selection of Xj from all candidate features. This test evaluates whether a variable's observed contribution is meaningfully higher than what would occur by random chance alone.

To make the findings more clinically accessible, Partial Dependence Plots (PDPs) were created for the highest-ranking predictors according to the multi-way importance results and the binomial test. These plots show the independent effect of each predictor on the estimated probability of death, while holding all other factors at their mean values. Such visualizations deliver a straightforward view of how individual variables relate to mortality risk.

Furthermore, to investigate possible nonlinear patterns and variable interactions, the average minimal depth of interactions was calculated inside the MixRFb framework.

Sensitivity analyses

Standalone Variable Predictive Analysis. The isolated forecasting ability of single variables was tested by constructing individual MixRFb models, each limited to a single variable ranked highly in the multi-way importance evaluation. Every variable was modeled separately to gauge its independent predictive strength.

Sensitivity Analysis with a Recurrent Neural Network. A Recurrent Neural Network (RNN) was employed as an additional robustness check to accommodate better the time-series nature of the repeated measurements using machine learning methods. The network was built with four input variables: age, gender, days spent in the ICU, and RDW. Training parameters included a batch size of 286 and five discrete time steps leading to a binary outcome.

Sensitivity Analysis with Generalized Linear Mixed-Effects Model. Another sensitivity test involved fitting a straightforward mixed-effects regression model.

Descriptive ROC Analysis. A standard descriptive Receiver Operating Characteristic (ROC) analysis was performed to compare the standalone discriminatory power of the classic SAPS score and RDW. Independent ROC curves were plotted for each of these two variables alone.

All analyses were performed in R version 4.3.1 [43], with support from the caret package [44].

Shiny application development

A dedicated Shiny web tool was built to provide clinicians with straightforward access to the MixRFb algorithm, allowing instant bedside calculation of ICU mortality risk.

Results and Discussion

This investigation presents the admission characteristics of 286 ICU patients, including 207 who survived their stay and 79 who did not. **Table 1** summarizes these features grouped by survival status. Clear statistical associations emerged for age, SAPS score, and diabetes. Survivors were typically younger, exhibited lower SAPS values, and had a lower prevalence of diabetes, pointing to milder acute illness severity upon entry. No meaningful links with survival were observed for gender, cardiovascular disease, respiratory disease, or the overall presence of comorbidities ($p > 0.05$).

Table 1. Baseline variables and repeated clinical measurements of ICU patients (N = 286).

Clinical feature	Overall (n = 286)	Deceased (n = 79)	Survivors (n = 207)	P-value	95% confidence interval	Odds ratio
Age at admission (years), median [IQR]	71.0 [61.0–78.0]	74.0 [66.0–79.0]	69.0 [59.0–78.0]	0.022	[1.00; 1.05]	1.03
Sex distribution, n (%)				0.496	[0.71; 2.25]	1.27
Male (reference group)	156 (62.4%)	39 (58.2%)	117 (63.9%)			
Female	94 (37.6%)	28 (41.8%)	66 (36.1%)			
Presence of diabetes mellitus, n (%)				0.046	[1.04; 3.42]	1.89
Absent (reference)	175 (70.0%)	40 (59.7%)	135 (73.8%)			
Present	75 (30.0%)	27 (40.3%)	48 (26.2%)			
Cardiovascular comorbidity, n (%)				0.138	[0.91; 2.84]	1.59
No history (reference)	122 (48.8%)	27 (40.3%)	95 (51.9%)			
Yes	128 (51.2%)	40 (59.7%)	88 (48.1%)			
Respiratory condition, n (%)				0.953	[0.44; 1.84]	0.92
No underlying disease (reference)	199 (79.6%)	54 (80.6%)	145 (79.2%)			
Yes	51 (20.4%)	13 (19.4%)	38 (20.8%)			
Severity score (SAPS), median [IQR]	40.0 [29.0; 50.0]	46.5 [39.0; 54.0]	36.0 [27.0; 46.0]	<0.001	[1.03; 1.08]	1.05
Any pre-existing comorbidity, n (%)				0.197	[0.85; 2.97]	1.56
None (reference)	85 (34.0%)	18 (26.9%)	67 (36.6%)			
At least one	165 (66.0%)	49 (73.1%)	116 (63.4%)			

Model performances

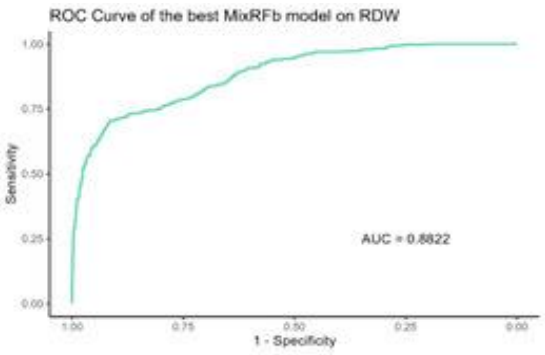
Training

• Model A (MixRFb that integrates RDW with additional patient factors): On the training data, this model delivered an AUC of 0.882 (95% CI: 0.860–0.904), demonstrating excellent ability to distinguish outcomes. The corresponding curve appears in **Figure 2a.**

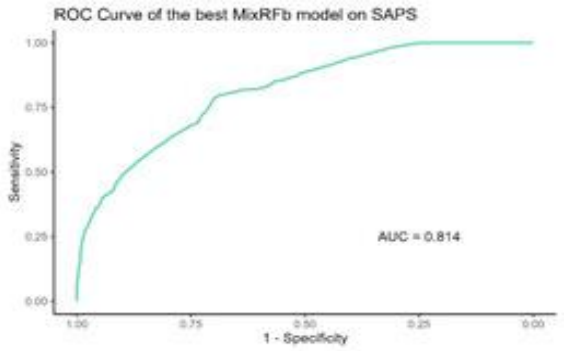
• Model B (MixRFb built around SAPS as the primary predictor): Training results were lower, yielding an AUC of 0.814 (95% CI: 0.790–0.838). This indicates that SAPS alone provides decent information, yet combining RDW with additional covariates, as in Model A, markedly improves predictive power.

Training performance for Model b is shown in **Figure 2b.**

• Model C (standard Random Forest relying on RDW as a predictor): Despite not using longitudinal repeated measures, the model still achieved a training AUC of 0.835 (95% CI: 0.812–0.858).



a)



b)

Figure 2. ROC curves for training performance of the MixRFb algorithm with RDW plus extra covariates (Panel a – Model a) and the MixRFb algorithm with SAPS (Panel b–Model b).

Validation

During bootstrap validation, Model A achieved a median F1-score of 0.76 (95% CI: 0.72–0.78) and a median AUC of 0.87 (95% CI: 0.85–0.88). These outcomes reinforced Model A's clear advantage in predictive performance compared with Models B and C under internal validation **(Table 2)**.

Table 2. AUC and F1 values obtained from bootstrap validation. The 95% confidence intervals are based on 1000 repeated iterations.

	F1	AUC
Model A	0.76 [0.72–0.78]	0.87 [0.85–0.88]
Model B	0.72 [0.69–0.74]	0.8 [0.79–0.83]
Model C	0.66 [0.72–0.77]	0.78 [0.8–0.81]

Variable contribution

The Variable Importance Plot (VIP) focused on the random forest section of the algorithm **(Figure 3)** underscores the key roles of age and RDW in forecasting death in the ICU **(Figure 3a)**.

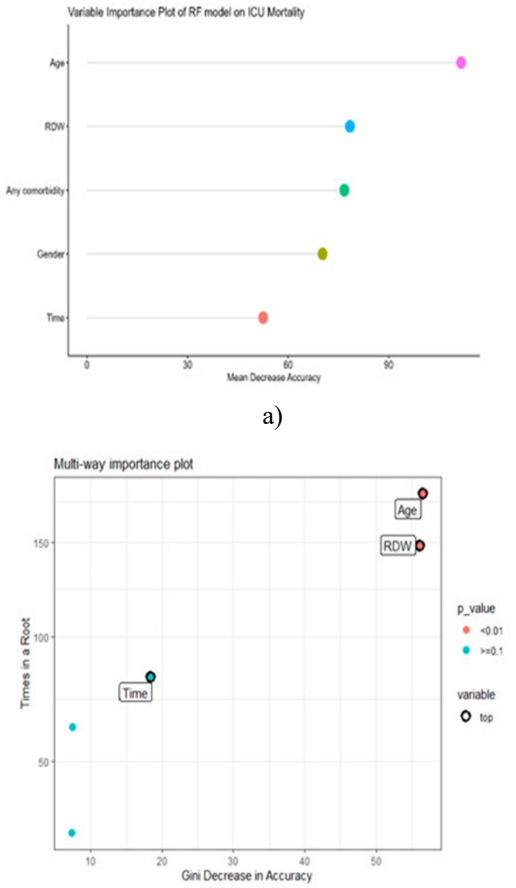


Figure 3. Variable importance plot illustrating mean decrease in accuracy for the RF classification component (Panel a). Multi-way importance plot (Panel b). This chart combines the average Gini reduction in accuracy — which quantifies each variable's overall usefulness to the model — with the frequency with which the variable is selected for root-node splits in the decision trees, thereby showing its effect on the most decisive early branching.

In addition, the multi-way importance plot confirmed that RDW and age were the dominant predictors, yielding the largest Gini reductions and appearing most frequently at root nodes. The time variable (days in ICU) also emerged as relevant, highlighting the algorithm's sensitivity to shifts in patient condition over the admission period **(Figure 3b)**.

Partial dependence plots **(Figure 4)** reveal the isolated impact of age, RDW, and time on estimated mortality probabilities. Age shows a nearly steady upward trend in risk as values rise (Panel a). RDW exhibits a steep rise in risk, especially beyond 16% (Panel b). Time produces largely flat predicted probabilities over days, consistent with stabilization among patients who remain longer in the ICU (Panel c).

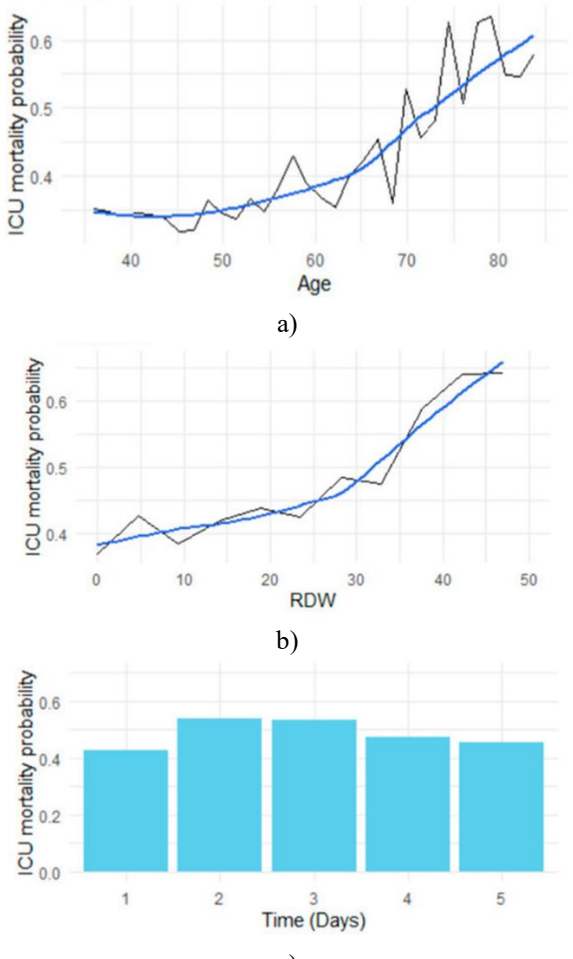


Figure 4. Partial Dependence Plots (PDPs) for central predictors of ICU mortality risk. (Panel a) Influence of age on predicted probability of death, (Panel b) Association between RDW (Red Cell Distribution Width) and predicted mortality, and (Panel c) Bar representation of predicted mortality risk across five successive days. Panels A and B present raw predictions in black and loess-smoothed curves in blue.

This evaluation pinpointed age and RDW as the most common and strongest interacting pair, with their combined effects typically arising near the tree roots.

Sensitivity analyses

Standalone Variable Predictive Analysis. To assess how well single variables could predict outcomes on their own, a MixRFb model was fitted using only the two top-ranked predictors identified by the multi-way importance analysis: age and RDW. When age was used in isolation, the model reached an AUC of 0.712 (95% CI: 0.687–0.734), indicating moderate discriminative capacity. By

comparison, RDW by itself produced a much lower AUC of 0.61 (95% CI: 0.58–0.63).

Sensitivity Analysis with a Recurrent Neural Network. The RNN approach delivered weaker overall results than the MixRFb algorithm, recording an AUC of 0.77 (95% CI: 0.75, 0.79) and an F1 score of 0.70 (95% CI: 0.68, 0.72). In direct contrast, the MixRFb model consistently outperformed it across all key evaluation measures.

Sensitivity Analysis with Mixed-Effects Model. The basic mixed-effects model yielded an AUC of 0.64 (95% CI: 0.4–0.74), which was noticeably lower than the performance of the MixRFb algorithm.

Descriptive ROC Analysis. Separate descriptive ROC curves were generated for the conventional SAPS score and for RDW used in isolation:

• Using SAPS alone: When evaluated independently, SAPS showed an AUC of 0.683 (95% CI: 0.655–0.711), pointing to relatively modest predictive strength.

• Using RDW alone: RDW exhibited the weakest standalone performance, with an AUC of 0.555 (95% CI: 0.527–0.583). This low value highlights its limited usefulness when applied without accounting for interactions or patterns involving other clinical factors.

Shiny app

The online predictive tool built around the top-performing algorithm (Model A) can be accessed at https://biostatlab24.shinyapps.io/MixRFbICU/ (accessed on 17 January 2025).

The MixRFb algorithm has been embedded in a Shiny application that prioritizes computational speed. Thanks to the inherently parallel nature of the random forest structure, the tool generates predictions very quickly.

This Shiny application allows users to estimate ICU mortality risk with the MixRFb algorithm (Model A). Clinicians can enter key patient information, including gender, age, RDW value, number of days in the ICU, and comorbidity status. Once the "Calculate" button is pressed, the app immediately presents the estimated probabilities of death and survival as a bar chart, supporting rapid risk assessment at the point of care. The layout is straightforward: input fields appear on the left side while the resulting prediction visualization is displayed on the right **(Figure 5)**.

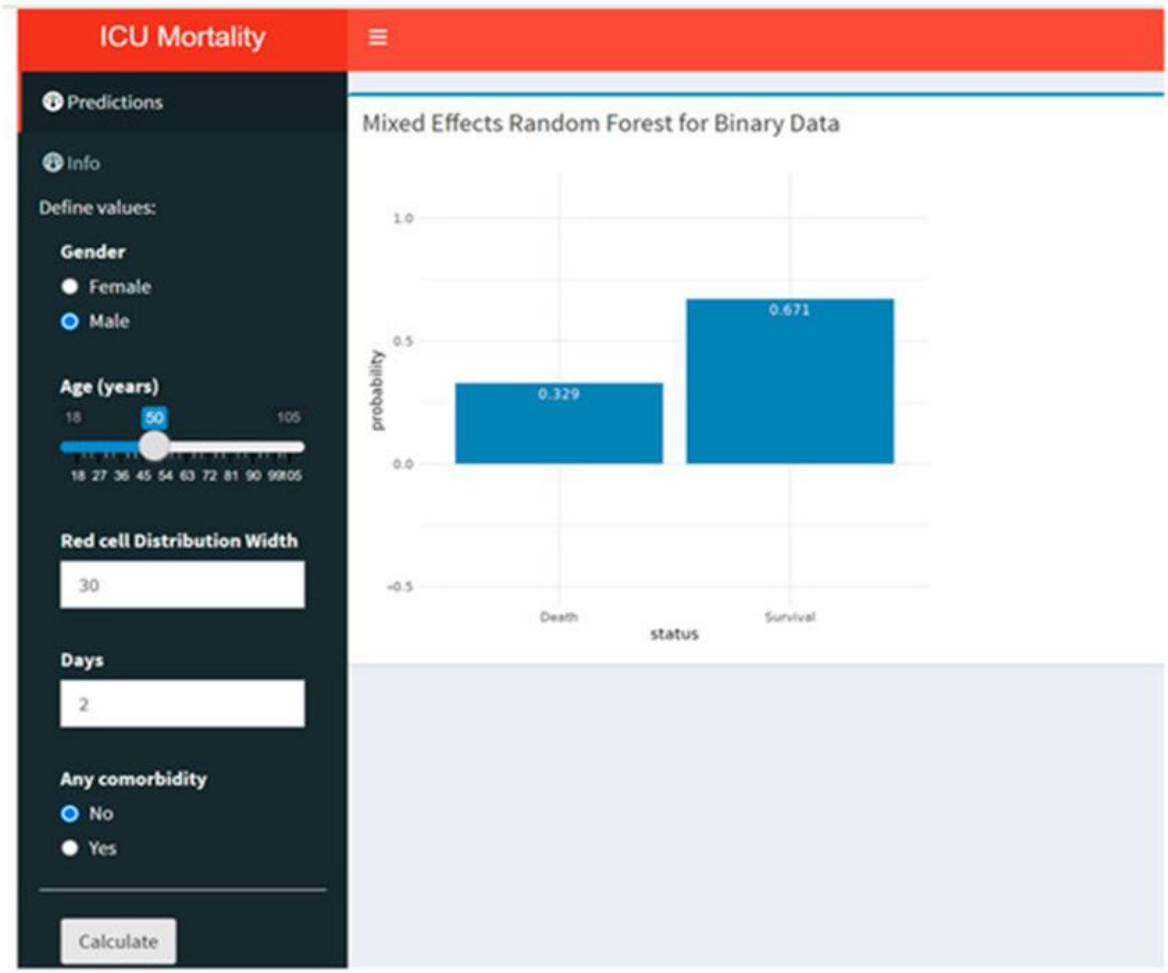


Figure 5. MixRFb online predictive tool.

The current work highlights the promise of the MixRFb algorithm for forecasting mortality among ICU patients by effectively incorporating repeated measurements over time. Its performance matches or surpasses that of several existing ICU mortality prediction tools reported in the literature [17, 45].

Compared with a conventional random forest model based solely on SAPS, the proposed MixRFb approach delivers a modest but meaningful gain in predictive accuracy. Importantly, it achieves this using readily obtainable and inexpensive variables such as RDW. While the SAPS score provides a solid initial snapshot of illness severity, it falls short in tracking a patient's condition as it evolves during the ICU admission. In contrast, changes in RDW — which the MixRFb framework can model longitudinally — mirror underlying processes such as systemic inflammation and oxidative stress, positioning it as a valuable and easily accessible marker for ongoing risk evaluation [46]. For example, two patients with nearly identical SAPS scores may follow very different clinical courses; an increasing RDW in one case might indicate emerging complications such as sepsis, an important signal that static SAPS scores overlook [47].

Furthermore, although SAPS is widely adopted, it depends on a broad array of physiological, laboratory, and clinical inputs, many of which demand considerable time and resources to collect [41]. This requirement can hinder its use for real-time decisions, including the swift identification of suitable candidates for clinical trials. By relying on only a small set of routinely available parameters, the MixRFb algorithm enables faster, more feasible patient stratification, thereby simplifying trial recruitment.

Embedding mixed-effects modeling inside the random forest structure brings notable benefits when dealing with typical ICU datasets. Such data naturally have a hierarchical structure because multiple observations are gathered from the same patient across different time points. The mixed-effects element explicitly accounts for the correlation among repeated measures within each individual, helping to reduce bias that would otherwise arise from this dependence [31]. As shown in both prior studies and the present findings, the MixRFb algorithm surpasses traditional mixed-effects regression in predictive performance. Its capacity to detect nonlinear patterns and complex higher-order interactions allows it to handle intricate data structures without the need to manually define interaction terms or apply data transformations [28]. The stronger results observed for Model A suggest that this approach is particularly effective at uncovering the multifaceted, nonlinear relationships present in ICU data, which is essential for reliable mortality forecasting [2].

By including longitudinal information such as length of ICU stay, the model successfully tracks shifts in patient status over time, overcoming a major shortcoming of traditional static scoring systems [36]. Moreover, the chosen covariates strike a practical balance between strong predictive value, clinical meaningfulness, and ease of use, supporting the creation of a robust yet widely deployable decision-support tool suitable for varied ICU environments [48].

Comorbidities and gender, even when they do not reach individual statistical significance, still play a valuable role in building a comprehensive risk profile by capturing differences in patient backgrounds and expected outcomes [40]. The independent performance of age alone demonstrates its substantial standalone contribution to forecasting ICU mortality. Both age and RDW emerged as the dominant predictors, consistent with their well-established importance in assessing patient frailty and overall physiological condition [23]. In particular, RDW serves as a strong clinical indicator, confirming its usefulness in intensive care settings and strengthening the overall predictive model [23]. In addition to its predictive value, the fact that RDW is routinely included in a standard complete blood count increases its real-world applicability. This widespread availability positions the MixRFb model as a convenient, practical solution for busy ICU environments, where timely, actionable information is essential [14, 23, 24]. Age helps gauge the baseline severity of illness at admission [35], whereas RDW serves as a dynamic indicator of systemic inflammation and physiological strain [26].

However, although age performs reasonably well when used alone, it fails to capture the combined interactions that the MixRFb model captures by integrating multiple factors, such as RDW, length of ICU stay, and comorbidities. Likewise, RDW alone has limited predictive power. Still, it gains considerable predictive power when paired with other clinical variables, acting as a marker of widespread inflammation that interacts with age and other factors. For instance, the strong interaction between age and RDW indicates that the joint impact of age-related frailty and ongoing inflammation has a greater effect on mortality risk than either factor considered separately [49].

The relatively modest standalone contribution of RDW does not diminish its clinical value; the partial dependence plot analysis shows a clear association between elevated RDW levels and increased mortality risk, consistent with its established role as an indicator of systemic inflammation and the severity of critical illness [50]. In a similar vein, the steady increase in mortality risk with advancing age underscores its central role in risk stratification. These results emphasize the advantage of employing combined predictive models rather than relying on single-variable assessments, particularly in critical care, where patient outcomes are shaped by multiple interconnected influences [51].

Furthermore, the MixRFb algorithm's superior performance relative to the widely used RNN model highlights its effectiveness in handling certain relational aspects of ICU data that the RNN struggled to capture. While RNNs are powerful, they may require substantially larger datasets or finer-grained temporal information to leverage their sequential processing strengths fully [52]. In comparison, the MixRFb model has been specifically adapted to the characteristics of the available ICU dataset. The MixRFb tool has the potential to support multicenter clinical trials in intensive care by providing a consistent method for selecting participants across different centers, thereby decreasing heterogeneity and strengthening the reliability of trial findings — especially in conditions such as acute kidney injury, where uniformity among patients is particularly important [9]. During periods of ICU strain, for example, in pandemics, the tool could help optimize scarce resources by identifying those patients most likely to benefit from aggressive interventions, promoting fair and efficient allocation while aiding ethically sound choices [19]. For individuals with longstanding illnesses such as COPD or heart failure, the tool enables more personalized management by forecasting trajectories based on serial clinical measurements, which may enhance recovery chances and shorten ICU length of stay [17]. Beyond the acute phase, its predictive output can inform post-ICU planning, guiding decisions about the intensity of follow-up care and helping to lower the chance of readmission; patients flagged as high-risk for post-discharge mortality, for instance, could be directed toward closer observation in intermediate care units or more intensive rehabilitation programs [21].

In line with the principles of precision medicine, the tool tailors therapeutic approaches to each patient's unique profile, thereby refining trial design and raising the chances of positive results in serious conditions such as sepsis.

In addition, embedding the algorithm in an intuitive Shiny web application promotes its uptake in daily clinical work, allowing healthcare teams to generate real-time predictions and base decisions on up-to-date risk estimates. This application helps close the divide between sophisticated machine learning methods and routine bedside practice, bringing advanced analytics within easy reach of frontline clinicians [18].

Issues related to data gathering or data quality could influence the reliability of predictions, so future work should focus on external validation of the algorithm to confirm its stability and broaden its usefulness across varied clinical environments and patient populations [45]. Since RDW proved to be a key predictor in the model, it is important to recognize its inherent constraints and possible sources of bias. Although RDW is a low-cost, routinely collected biomarker, differences between laboratories may compromise its uniformity and limit the extent to which results generalize [53]. In heterogeneous ICU environments, variations in laboratory methods, data recording practices, and the completeness of patient records can pose challenges to the model's transportability [50]. For example, RDW results may differ depending on the specific analyzers and techniques employed. Establishing uniform protocols for RDW measurement would help strengthen the model's consistency. Looking ahead, subsequent studies could integrate local calibration steps to enable the algorithm to adapt to site-specific patterns without losing its core predictive power. Despite these challenges, the ease of access to RDW highlights its practical usefulness. It marks it as a promising element for advancing predictive tools, while also pointing to clear directions for continued improvement and investigation [54].

The MixRFb algorithm runs efficiently enough to deliver instant predictions in busy ICU settings, even on ordinary computing equipment. Its scalability is enhanced by the accompanying web application, which simplifies rollout across different environments. Nevertheless, an effective rollout will require immediate access to live clinical information and seamless integration with existing electronic health record platforms. Cloud computing options could further boost both scalability and availability, especially in busy or under-resourced intensive care units.

Although the MixRFb algorithm shows better results than SAPS, the current comparison is restricted to this one scoring system alone. Several other well-known ICU prediction tools, including the Acute Physiology and Chronic Health Evaluation (APACHE) [55] and the Mortality Probability Models (MPM) [12], are also commonly used to estimate mortality risk in critically ill patients. Like SAPS, these systems mainly depend on information collected at admission and make little use of repeated measurements over time. Additional research should test the MixRFb algorithm directly against APACHE and MPM to allow a broader evaluation and strengthen confidence in its value across a wider range of ICU environments.

Furthermore, although bootstrap-based internal validation indicates strong internal stability, testing the model on completely separate external datasets remains essential to assess its performance across varied ICU populations and settings. Dedicated external validation studies will be required to verify the algorithm's consistency and usefulness outside the original study site.

In addition, while the algorithm generates valuable risk estimates, it is not intended to substitute for clinical expertise. Instead, it should function as a supportive aid that helps clinicians reach better-informed choices. Ethical principles must always direct their application, ensuring the tool enhances rather than supplants the judgment of seasoned medical professionals.

Conclusion

This study demonstrates the promise of the MixRFb algorithm for improving mortality prediction in intensive care by leveraging time-varying data, particularly the inexpensive and widely available RDW marker. The algorithm surpassed conventional approaches such as the SAPS score, proving more effective at handling the complex nature of ICU data. The creation of a Shiny web application turns this method into a practical bedside resource, supporting better participant selection for clinical trials, more efficient resource allocation during periods of ICU overload, and more individualized care plans. Upcoming investigations should prioritize external validation and broader testing of the algorithm in different clinical contexts to further advance both patient management in intensive care and the design of future clinical trials.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: The data collection was approved by the local Ethics Committee (CE AVEC) under protocol number 160699, with approval date 14 July 2016.
Informed consent was obtained from all subjects involved in the study.

References

1. Delaney A, Angus DC, Bellomo R, Cameron P, Cooper DJ, Finfer S, et al. Bench-to-bedside review: the evaluation of complex interventions in critical care. Crit Care. 2008;12(2):210.
2. Granholm A, Alhazzani W, Derde LPG, Angus DC, Zampieri FG, Hammond NE, et al. Randomised clinical trials in critical care: past, present and future. Intensive Care Med. 2022;48(2):164-78.
3. François B, Clavel M, Vignon P, Laterre PF. Perspective on optimizing clinical trials in critical care: how to puzzle out recurrent failures. J Intensive Care. 2016;4:67.
4. Ali NA, Wolf KM, Hammersley J, Hoffmann SP, O'Brien JM, Phillips GS, et al. A cluster-randomized trial of intensivist staffing. Am J Respir Crit Care Med. 2011;184(7):803-8.
5. Luce JM, Cook DJ, Martin TR, Angus DC, Boushey HA, Curtis JR, et al. The ethical conduct of clinical research involving critically ill patients in the United States and Canada: principles and recommendations. Am J Respir Crit Care Med. 2004;170(12):1375-84.
6. Yamga E, Mantena S, Rosen D, Bucholz EM, Yeh RW, Celi LA, et al. Optimized risk score to predict mortality in patients with cardiogenic shock in the cardiac intensive care unit. J Am Heart Assoc. 2023;12(13):e029232.
7. Boussina A, Shashikumar SP, Malhotra A, Owens RL, El-Kareh R, Longhurst CA, et al. Impact of a deep learning sepsis prediction model on quality of care and survival. NPJ Digit Med. 2024;7(1):14.
8. Villar J, Ferrando C, Tusman G, Berra L, Rodríguez-Suárez P, Suárez-Sipmann F. Unsuccessful and successful clinical trials in acute respiratory distress syndrome: addressing physiology-based gaps. Front Physiol. 2021;12:774025.
9. Koshiaris C, Archer L, Lay-Flurrie S, Snell KI, Riley RD, Stevens R, et al. Predicting the risk of acute kidney injury in primary care: derivation and validation of STRATIFY-AKI. Br J Gen Pract. 2023;73(734):e605-14.
10. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med. 1981;9(8):591-7.
11. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. Crit Care Med. 1984;12(11):975-7.
12. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. JAMA. 1993;270(20):2478-86.
13. Zhai Q, Lin Z, Ge H, Liang Y, Li N, Ma Q, et al. Using machine learning tools to predict outcomes for emergency department intensive care unit patients. Sci Rep. 2020;10(1):20919.
14. Tian Y, Yao Y, Zhou J, Diao X, Chen H, Cai K, et al. Dynamic APACHE II score to predict the outcome of intensive care unit patients. Front Med (Lausanne). 2022;8:744907.
15. Atallah L, Nabian M, Brochini L, Amelung PJ. Machine learning for benchmarking critical care outcomes. Healthc Inform Res. 2023;29(4):301-14.
16. Sikora A, Zhang T, Murphy DJ, Smith SE, Murray B, Kamaleswaran R, et al. Machine learning vs. traditional regression analysis for fluid overload prediction in the ICU. Sci Rep. 2023;13(1):19654.
17. Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. Sci Rep. 2019;9(1):11862.
18. Chen Z, Li T, Guo S, Zeng D, Wang K. Machine learning-based in-hospital mortality risk prediction tool for intensive care unit patients with heart failure. Front Cardiovasc Med. 2023;10:1119699.
19. Hong WS, Rudas A, Bell EJ, Chiang JN. Association of red blood cell distribution width with hospital

admission and in-hospital mortality across all-cause adult emergency department visits. JAMIA Open. 2023;6(2):ooad053.
20. Fontana V, Spadaro S, Villois P, Righy Shinotsuka C, Fogagnolo A, Nobile L, et al. Can red blood cell distribution width predict outcome after cardiac arrest? Minerva Anestesiol. 2018;84(6):693-702.
21. Yonemoto S, Hamano T, Fujii N, Shimada K, Yamaguchi S, Matsumoto A, et al. Red cell distribution width and renal outcome in patients with non-dialysis-dependent chronic kidney disease. PLoS One. 2018;13(6):e0198825.
22. Valko L, Baglyas S, Podmaniczky E, Prohaszka Z, Gal J, Lorx A. Exploring red cell distribution width as a biomarker for treatment efficacy in home mechanical ventilation. BMC Pulm Med. 2022;22(1):115.
23. Fogagnolo A, Spadaro S, Taccone FS, Ragazzi R, Romanello A, Fanni A, et al. The prognostic role of red blood cell distribution width in transfused and non-transfused critically ill patients. Minerva Anestesiol. 2019;85(11):1159-67.
24. Valenti AC, Vitolo M, Imberti JF, Malavasi VL, Boriani G. Red cell distribution width: a routinely available biomarker with important clinical implications in patients with atrial fibrillation. Curr Pharm Des. 2021;27(37):3901-12.
25. She Y, Li Y, Chen S, Chen Y, Zhou L. Red blood cell distribution width predicts in-hospital mortality in patients with a primary diagnosis of seizures in the ICU: a retrospective database study. Neurol Sci. 2022;43(1):499-506.
26. Horne BD, Anderson JL, Muhlestein JB, Ridker PM, Paynter NP. Complete blood count risk score and its components, including RDW, are associated with mortality in the JUPITER trial. Eur J Prev Cardiol. 2015;22(4):519-26.
27. Lee J. Patient-specific predictive modeling using random forests: an observational study for the critically ill. JMIR Med Inform. 2017;5(1):e3.
28. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. Brief Bioinform. 2023;24(1):bbad002.
29. Dankl D, Rezar R, Mamandipoor B, Zhou Z, Wernly S, Wernly B, et al. Red cell distribution width is independently associated with mortality in sepsis. Med Princ Pract. 2022;31(2):187-94.
30. Hu S, Wang YG, Drovandi C, Cao T. Predictions of machine learning with mixed-effects in analyzing longitudinal data under model misspecification. Stat Methods Appl. 2023;32(3):681-711.
31. Chen A, Zhao Z, Hou W, Singer AJ, Li H, Duong TQ. Time-to-death longitudinal characterization of clinical variables and longitudinal prediction of mortality in COVID-19 patients: a two-center study. Front Med (Lausanne). 2021;8:661940.
32. Wang J, Gamazon ER, Pierce BL, Stranger BE, Im HK, Gibbons RD, et al. Imputing gene expression in uncollected tissues within and beyond GTEx. Am J Hum Genet. 2016;98(4):697-708.
33. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18-22.
34. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1-67.
35. Gonçalves-Pereira J, Oliveira A, Vieira T, Rodrigues AR, Pinto MJ, Pipa S, et al. Critically ill patient mortality by age: long-term follow-up (CIMbA-LT). Ann Intensive Care. 2023;13(1):7.
36. Forte JC, van der Horst ICC. Comorbidities and medical history essential for mortality prediction in critically ill patients. Lancet Digit Health. 2019;1(2):e48-9.
37. Siegelaar SE, Hickmann M, Hoekstra JB, Holleman F, DeVries JH. The effect of diabetes on mortality in critically ill patients: a systematic review and meta-analysis. Crit Care. 2011;15(5):R205.
38. Rocha B, Cunha G, Maltes S, Moura AN, Coelho F, Torres J, et al. Cardiovascular disease in an intensive care unit: patterns of an often fatal omen. Eur Heart J. 2021;42(Suppl 1):ehab724.1520.
39. Grangier B, Vacheron CH, De Marignan D, Casalegno JS, Couray-Targe S, Bestion A, et al. Comparison of mortality and outcomes of four respiratory viruses in the intensive care unit: a multicenter retrospective study. Sci Rep. 2024;14(1):6690.
40. Merdji H, Long MT, Ostermann M, Herridge M, Myatra SN, De Rosa S, et al. Sex and gender differences in intensive care medicine. Intensive Care Med. 2023;49(10):1155-67.
41. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA. 1993;270(24):2957-63.
42. Lin Z, Lawrence WR, Huang Y, Lin Q, Gao Y. Classifying depression using blood biomarkers: a large population study. J Psychiatr Res. 2021;140:364-72.
43. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2023.
44. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1-26.
45. Li F, Xin H, Zhang J, Fu M, Zhou J, Lian Z. Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database. BMJ Open. 2021;11(7):e044779.

46. Moreno-Torres V, Royuela A, Múñez-Rubio E, Gutierrez-Rojas Á, Mills-Sánchez P, Ortega A, et al. Red blood cell distribution width as prognostic factor in sepsis: a new use for a classical parameter. J Crit Care. 2022;71:154069.
47. Duggal A, Scheraga R, Sacha GL, Wang X, Huang S, Krishnan S, et al. Forecasting disease trajectories in critical illness: comparison of probabilistic dynamic systems to static models to predict patient status in the intensive care unit. BMJ Open. 2024;14(2):e079243.
48. Vallet H, Guidet B, Boumendil A, de Lange DW, Leaver S, Szczeklik W, et al. The impact of age-related syndromes on ICU process and outcomes in very old patients. Ann Intensive Care. 2023;13(1):68.
49. Said AS, Spinella PC, Hartman ME, Steffen KM, Jackups R, Holubkov R, et al. RBC distribution width: biomarker for red cell dysfunction and critical illness outcome? Pediatr Crit Care Med. 2017;18(2):134-42.
50. Peng S, Li W, Ke W. Association between red blood cell distribution width and all-cause mortality in unselected critically ill patients: analysis of the MIMIC-III database. Front Med (Lausanne). 2023;10:1152058.
51. Stewart J, Bradley J, Smith S, McPeake J, Walsh T, Haines K, et al. Do critical illness survivors with multimorbidity need a different model of care? Crit Care. 2023;27(1):485.
52. Waqas M, Humphries UW. A critical review of RNN and LSTM variants in hydrological time series predictions. MethodsX. 2024;13:102946.
53. Li N, Zhou H, Tang Q. Red blood cell distribution width: a novel predictive indicator for cardiovascular and cerebrovascular diseases. Dis Markers. 2017;2017:7089493.
54. Zhang Z, Chew GM, Shikuma CM, Gangcuangco LM, Souza SA, Shiramizu B, et al. Red blood cell distribution width as an easily measurable biomarker of persistent inflammation and T cell dysregulation in antiretrovirally treated HIV-infected adults. HIV Clin Trials. 2018;19(5):172-6.
55. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818-29.