# Predicting sepsis prognosis with machine learning models trained on MIMIC-IV

**Felipe Alexandre P. Miranda**[1] , **Fábio Agostini A. Gomes**[1,2] , **Sarah Negreiros de Carvalho**[1]

[1]Aeronautics Institute of Technology (ITA) – Brazil

[2]Irmandade da Santa Casa de Misericórdia de São Paulo – Brazil

felipe.miranda.8757@ga.ita.br

***Abstract.*** *Sepsis remains one of the leading causes of mortality in intensive care units (ICUs), making early prognosis prediction essential for improving clinical outcomes. This study develops and compares five machine learning models—Logistic Regression, Gradient Boosting, XGBoost, LightGBM, and Multi-Layer Perceptron—to predict 30-day mortality in 6,965 sepsis patients from the MIMIC-IV database. Our key contribution is demonstrating that a model-agnostic feature selection approach reduces the variable set from 103 to 20 clinical parameters while maintaining equivalent predictive performance (AUC ¿ 0.87 across all models), significantly improving computational efficiency and clinical interpretability. Systematic hyperparameter optimization with algorithm-specific class-imbalance strategies revealed that ensemble methods (XGBoost, LightGBM, MLP) achieved validation AUC values above 0.90. The identified minimal feature set, dominated by hemodynamic and metabolic markers, provides actionable clinical information while establishing reproducible benchmarks for sepsis mortality prediction, demonstrating the potential of machine learning as a practical support tool for ICU decision-making.*
*Sepsis Mortality Prediction Machine Learning in Healthcare Feature Selection.*

## 1. Introduction

Sepsis is a critical clinical condition characterized by an uncontrolled systemic inflammatory response triggered by infection. It often progresses to multiple-organ failure and remains one of the leading causes of mortality in intensive care units (ICUs), with substantial socioeconomic impact due to prolonged, intensive treatment [Arbous et al. 2024]. With global sepsis incidence exceeding 48 million cases annually and mortality rates ranging from 15-30% in ICUs [Singer et al. 2016], the development of accurate prognostic tools represents a critical clinical priority.

Early and accurate prognosis is essential to reduce sepsis mortality, yet traditional scores like the Sequential Organ Failure Assessment (SOFA) and Acute Physiology and Chronic Health Evaluation (APACHE II) are limited by predetermined linear weights that fail to capture complex pathophysiological interactions [Singer et al. 2016]. Machine learning (ML) algorithms offer a powerful alternative by modeling non-linear patterns, with studies demonstrating significant performance gains over conventional methods, including AUC improvements of 0.05 to 0.15[Han et al. 2025, Meng et al. 2022,

Yu et al. 2025]. However, a critical gap remains in developing clinically feasible models that balance this predictive accuracy with the practical demands of ICU settings, such as computational efficiency, interpretability, and real-time feature availability.

The MIMIC-IV (Medical Information Mart for Intensive Care IV) database provides high-quality, real-world ICU data—encompassing vital signs, laboratory results, and clinical interventions from over 300,000 ICU admissions—supporting robust model development and validation [Johnson et al. 2024]. Prior work with MIMIC-IV or similar cohorts has demonstrated strong performance from ensemble tree-based methods (XGBoost, LightGBM, and neural networks) when predicting sepsis-related outcomes, with reported AUC values typically ranging from 0.80 to 0.90 [Bao et al. 2023, Shan et al. 2024]. However, these studies often utilize extensive feature sets (¿100 variables) that may not be readily available in all clinical settings, and lack systematic comparisons of algorithm-specific optimization strategies for the inherent class-imbalance challenges in mortality prediction.

This study addresses these critical gaps by developing a comprehensive machine learning framework specifically designed for clinical implementation in sepsis mortality prediction. Our primary contributions are fourfold: (1) We demonstrate that a model-agnostic feature selection methodology can identify a minimal set of 20 clinical variables that maintain predictive performance equivalent to comprehensive datasets (103 variables) across five different algorithm families, achieving AUC values above 0.87 while reducing computational requirements and improving clinical interpretability; (2) We provide a systematic evaluation of algorithm-specific class-imbalance mitigation strategies in the sepsis prediction context, establishing optimal hyperparameter configurations that enhance minority class detection (mortality prediction) while maintaining overall model stability and generalizability; (3) We reveal consistent feature importance patterns across diverse algorithmic approaches, with hemodynamic instability markers and metabolic dysfunction indicators consistently emerging as the most predictive variables, providing actionable clinical insights that align with established sepsis pathophysiology; and (4) We establish reproducible benchmarks for 30-day mortality prediction in sepsis patients using MIMIC-IV data, with comprehensive performance evaluation that includes not only discrimination metrics but also calibration analysis and clinical interpretability assessments. To achieve these objectives, five classification algorithms—Logistic Regression (LR), Gradient Boosting (GB), XGBoost, LightGBM and Multi-Layer Perceptron (MLP)—are systematically trained and optimized using rigorous hyperparameter tuning with algorithm-specific strategies for handling class imbalance. Model-specific feature-selection techniques are then applied to identify the minimal variable set required to maintain predictive performance, to develop clinically-feasible models that can be readily implemented in real-world ICU settings while maintaining transparency and interpretability for clinical decision-making.

## 2. Methodology

### 2.1. Dataset Description

This study utilizes the MIMIC-IV database, which contains detailed clinical information on patients admitted to intensive care units at the Beth Israel Deaconess Medical Center in the United States, covering admissions between 2008 and 2019. This database is

widely recognized for the quality and comprehensiveness of its data, enabling robust investigations into various clinical and technological aspects of intensive care medicine, and is extensively used in both clinical and technological research in hospital settings [Johnson et al. 2024]. Table 1 presents the data used from the MIMIC-IV database.

**Table 1. Data files and directories.**

| Information type | CSV file | Folder |
|---|---|---|
| Patients and admissions | `patients.csv`, `admissions.csv` | `hosp` |
| Diagnoses | `diagnoses_icd.csv` | `hosp` |
| Laboratory tests | `labevents.csv` | `hosp` |
| Vital signs and clinical events | `chartevents.csv` | `icu` |
| Infections | `microbiologyevents.csv` | `hosp` |
| ICU stay | `icustays.csv` | `icu` |

For this study, 6,965 patients diagnosed with general and common types of sepsis were selected using International Classification of Diseases (ICD) codes: A419 (unspecified sepsis), 0389 (unspecified sepsis – ICD-9), R6521 (severe sepsis with septic shock), R652 (severe sepsis at risk of shock – ICD-10), 99592 (severe sepsis at risk of shock – ICD-9), and 4773 (unspecified septicemia). These categories were chosen as they represent the most frequent and comprehensive sepsis diagnoses within the studied population, allowing for a generalized and clinically relevant approach to predictive model development. After splitting the dataset into training (80%) and test (20%) sets, the class distributions were preserved through stratification. The training set contained 1,484 samples from class 1 (deaths) and 4,088 from class 0 (survivors), while the test set included 371 deaths and 1,022 survivors. A comprehensive exploratory data analysis was then conducted to examine variable distributions, identify outliers, and investigate potential relationships between predictors and clinical outcomes.

### 2.2. Preprocessing and Initial Feature Selection

Preprocessing steps addressed missing values, normalized continuous variables, and one-hot–encoded categorical features. A 30-day mortality window was adopted because most sepsis-related deaths occur within the first month, limiting late-stage confounders and matching common practice in the literature [Karakike et al. 2019, Al Omar et al. 2024].

Laboratory tests were retained only if ordered for at least 80% of patients; vital-sign variables met a higher 90% threshold, reflecting their near-continuous monitoring. These cut-offs balanced clinical coverage against noise from rarely measured parameters. For each retained variable, the minimum and maximum values recorded during the hospital stay were kept. These extremes capture clinically relevant fluctuations, flagging acute deterioration while simultaneously reducing dimensionality. A correlation screen then removed predictors weakly linked to 30-day mortality ($|r| < 0.03$) and those highly collinear with others ($r > 0.90$), minimizing redundancy and potential bias. Because 59.85% of patients had complete records, only fully observed cases were kept, yielding 6 965 patients and 103 features. Continuous variables were min–max scaled to $(0, 1)$, and nominal categories (e.g., `admission_type`) were one-hot encoded before model training.

### 2.3. Exploratory Data Analysis

Due to the large number of available variables, a systematic analysis approach was necessary. The parameters were organized into ten physiological groups, following the methodology proposed in [Shan et al. 2024], with one representative variable selected from each group, as shown in Table 2.

**Table 2. Representative features from each physiological group.**

| Physiological Group | Representative Feature |
|---|---|
| Heart Indicators | heart_rate_min |
| Blood Pressure Indicators | nbpd_min |
| Respiratory Indicators | resp_rate_min |
| Temperature Indicators | temperature_f_min |
| Blood Sugar Indicators | glucose_vitalsign_mean |
| Blood Gas Indicators | ph_max |
| Blood Indicators | wbc_min |
| Coagulation Function Indicators | inr_min |
| Glasgow Coma Scale (GCS) | motor_response_max |

The distributions of these representative variables revealed clinically relevant patterns (Figure 1). Markers of severe instability, such as near-zero minimum heart and respiratory rates, were more frequent in the non-survivor group, suggesting cardiac arrest events. Similarly, lower minimum blood pressure ($nbpd\_min$) and Glasgow Coma Scale scores ($motor\_response\_max$) were associated with higher mortality, consistent with septic shock and encephalopathy. As expected, mortality risk increased progressively with age, exceeding 30% in patients over 80 years old (Figure 2).

Figure 3 shows the distribution of the categorical variables "sex", "race", and "admission type" across the classes. "Sex" showed a balanced distribution, with no significant impact on mortality. "Race" revealed a slight increase in mortality among patients classified as "Other"; however, this observation should be interpreted with caution due to the limited sample size. "Admission type" showed similar rates across categories, except for the "Other" class, which had higher mortality but also a low relative frequency.

Additionally, the correlations between these representative variables and the outcome variable were analyzed, as shown in the heatmap in Figure 4. The variable $heart\_rate\_min$ had the highest negative correlation with 30-day mortality (-0.32), indicating that extremely low heart rate values are associated with a higher risk of death. Age showed a positive correlation of 0.13, consistent with the increased mortality in older age groups. The variable $inr\_min$ also showed a moderate positive correlation (0.18), suggesting that coagulation disorders may be related to worse outcomes. Other variables, such as $wbc\_min$ (0.065) and $motor\_response\_max$ (-0.19), showed weaker correlations but still suggested some clinical impact. Variables like $glucose\_min$, $ph\_max$, $temperature\_f\_min$, and $nbpd\_min$ showed very low correlations with mortality, suggesting they have low predictive power when considered alone.

Among the clinical variables themselves, a positive correlation was observed between $heart\_rate\_min$ and $resp\_rate\_min$ (0.39), reflecting the physiological pattern of

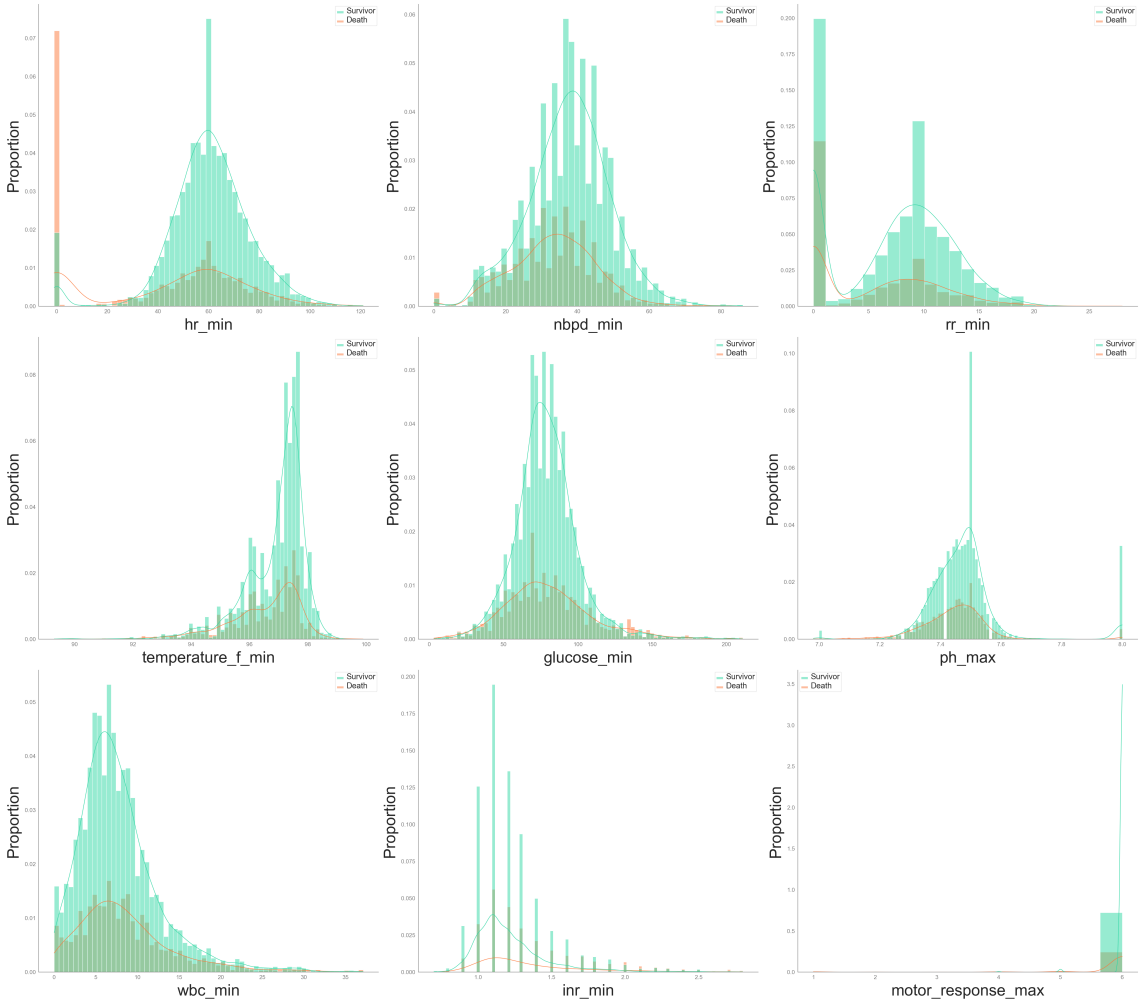


**Figure 1. Distribution of selected clinical variables stratified by 30-day mortality. Each subplot compares the normalized frequency (proportion) of survivors and deaths for a specific variable, highlighting potential differences in vital signs and laboratory measurements.**

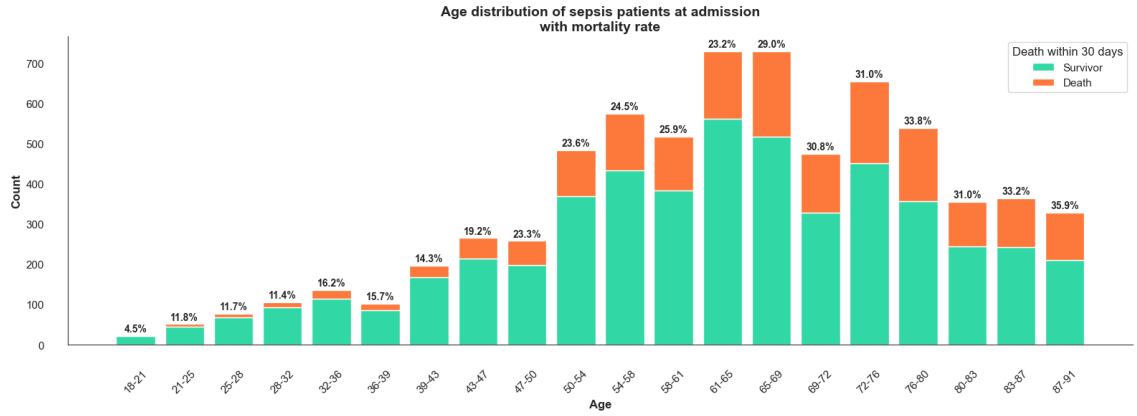


**Figure 2. Patient age as a mortality risk factor. The left panel shows the cohort's age distribution. The right panel plots mortality rate per age bracket, revealing a clear trend where mortality risk increases significantly with age.**

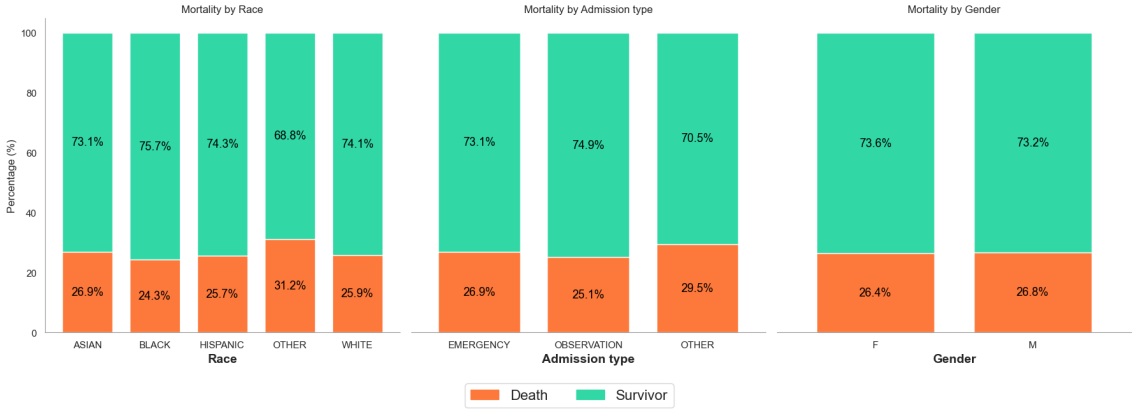


**Figure 3. Distribution of demographic variables by mortality outcome. The balanced proportions between survivors and deceased across most categories suggest these variables are weak individual predictors of mortality.**

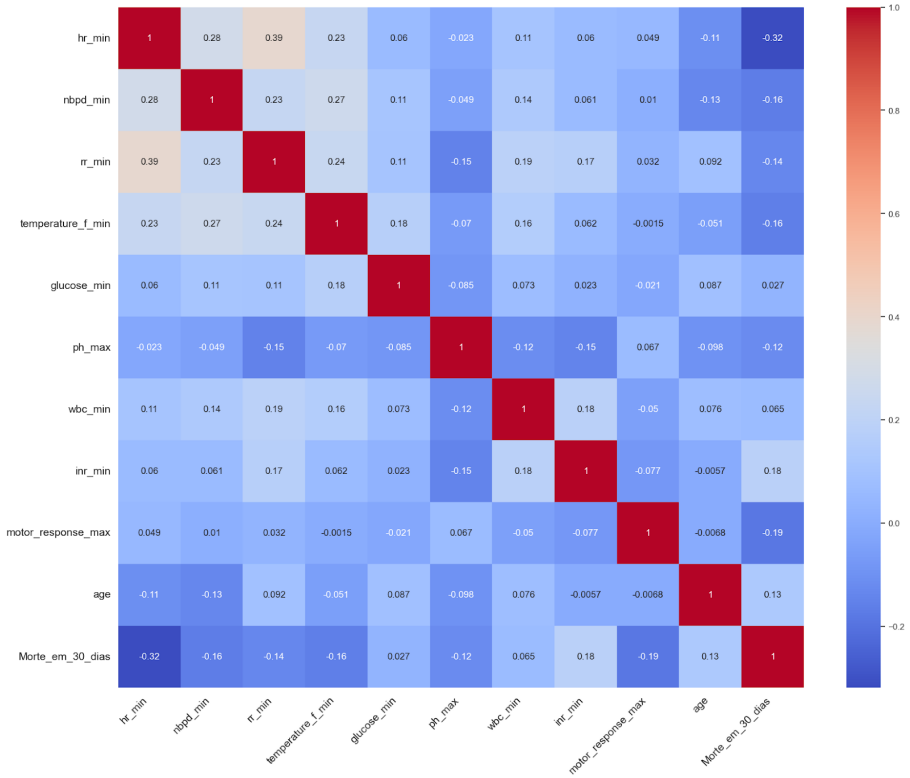


**Figure 4. Heatmap showing the correlation between representative variables and the outcome label (30-day mortality) in septic patients.**

compensatory response in septic conditions. $ph\_max$ showed a negative correlation with $resp\_rate\_min$ (-0.15), possibly reflecting compensatory acid-base disorders. Regarding the extreme values, a conservative approach was applied. The criterion of five standard deviations from the mean was adopted, minimizing the risk of excluding valid observations. For continuous variables with less than 10% outliers, extreme values were replaced by the mode, assuming these were recurring measurement errors. Finally, the less informative categorical variables were kept in the dataset for later evaluation of their relative importance in predictive models.

## 3. Models Architecture

**Table 3. Optimal hyperparameters after 30 Optuna trials and the class-imbalance strategy adopted by each model.**

| Model | Key hyperparameters (optimum) | Imbalance strategy |
|---|---|---|
| LR | solver=`liblinear`<br>$C = 9.65$<br>penalty=L2<br>max_iter=1000 | class_weight = 'balanced' |
| GB | $n_{\text{estimators}} = 300$<br>depth=3, $\eta = 0.01$<br>subsample=0.80<br>max_feat=0.80; min_leaf=30 | class_weight = 'balanced' |
| XGBoost | $n_{\text{estimators}} = 300$<br>depth=5, $\eta = 0.001$<br>subsample=0.606; colsample=0.634<br>$\gamma = 0.58$; $\alpha = 0.42$; $\lambda = 2.61$; min_child_wt=7 | scale_pos_ weight = 2.75 |
| LightGBM | $n_{\text{estimators}} = 1000$<br>depth=6, $num\_leaves = 20$<br>bagging_frac=0.894; feature_frac=0.791<br>$\lambda_1 = 0.390$; $\lambda_2 = 0.003$; min_child=30 | is_unbalance = True |
| MLP | layers=(192,64)<br>dropout=0.40; $L2 = 5.6 \times 10^{-5}$<br>$\eta = 0.0052$; batch_size=64<br>epochs=50[†]; early-stop | class-weighted binary cross-entropy |

[†] Training stops earlier if the validation loss does not improve for five consecutive epochs.

All experiments use a shared pre-processing pipeline implemented in `scikit-learn`. Numerical variables are imputed with the median, categorical variables are one-hot encoded, and standard scaling is applied only to LR and MLP—the two algorithms whose optimization procedures are sensitive to feature magnitude. Given the target imbalance (approximately $26\%$ of records correspond to deaths), each algorithm incorporates class-reweighting mechanisms: inverse-frequency weights for LR and GB, `scale_pos_weight` in XGBoost, the `is_unbalance` flag in LightGBM, and class-weighted binary cross-entropy in the MLP.

Hyperparameters for the five classifiers are optimized using random search via the Optuna framework, with 30 trials of five-fold stratified cross-validation maximizing F1-score. For boosting models, early stopping with 50 rounds of patience prevents overfitting. Following optimization, models are retrained on the complete training set before test evaluation. The search space and the best values obtained for each classifier are summarized in Table 3, together with the specific strategy adopted to mitigate class imbalance. To ensure uniform comparison, feature selection is performed once using model-agnostic SHAP analysis on a baseline XGBoost. The 20 most influential variables form a reduced feature set used by all classifiers, simplifying interpretation and reducing computational cost while enabling consistent assessment of dimensionality reduction effects. Final performance is reported on the independent test set using AUROC, accuracy, precision, re-

call, specificity, and F1-score, following the guidelines of [Huayanay et al. 2025].

## 4. Results and Discussion

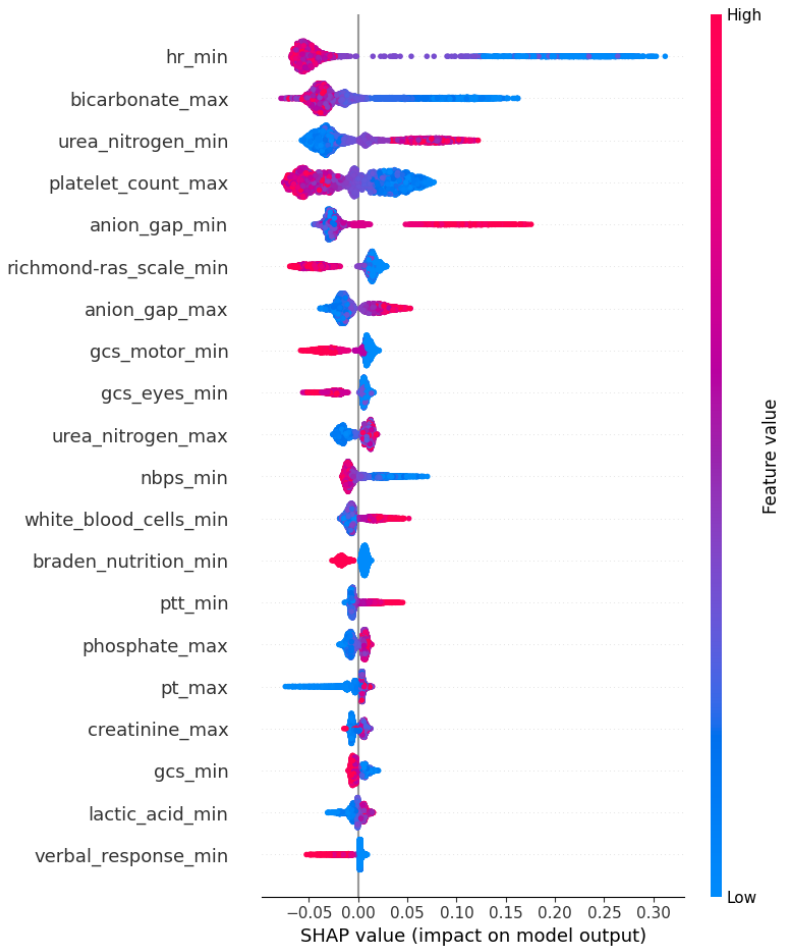


**Figure 5. Top 20 features selected for the XGBoost model using SHAP values, indicating each feature's contribution to the model output.**

All results presented in this section were obtained with the hyperparameters selected by the 30-trial Optuna random search (5-fold stratified cross-validation) described in Section 3. During tuning, the minority class was up-weighted inside each algorithm so that every model was trained under the same 26% mortality imbalance. After tuning, the five classifiers—LightGBM, XGBoost, Gradient Boosting, Logistic Regression and the MLP—underwent a feature-selection step. The procedure mirrored the workflow adopted for each model family: LightGBM and GB used their internal split-gain importance, XGBoost and the MLP relied on SHAP values, and LR used recursive feature elimination (RFE) based on the coefficient magnitude. The twenty highest-ranked variables in each case were retained for a reduced version of the model. Figures 5 and 6 illustrate the selected features for XGBoost and LightGBM, respectively; note the dominance of haemodynamic and metabolic markers (mean arterial pressure, lactate, heart-rate variability), in agreement with the exploratory patterns observed in Section 2.3.

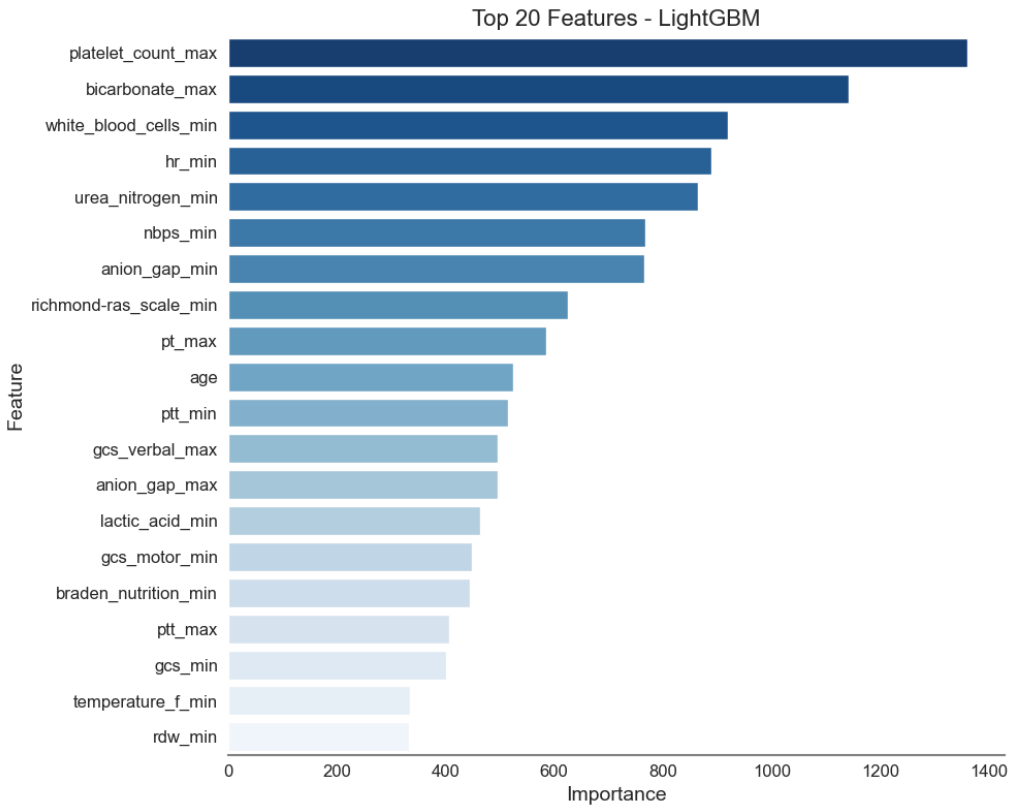


**Figure 6. Top 20 features selected for the LightGBM model using `feature_importance`.**

The practical value of this feature analysis is twofold. First, the consistent selection of markers well-understood in sepsis pathophysiology (e.g., lactate, pH, heart rate variability, and blood pressure) builds clinical trust. Clinicians are more likely to adopt a model whose reasoning aligns with established medical knowledge. Second, it provides a clear pathway for implementation: a clinical decision support tool could focus on tracking these 20 key variables. Instead of a generic alert, the system could specify which patient parameters are driving the high-risk score, prompting targeted actions like fluid resuscitation for hypotension or further investigation into metabolic acidosis.

Table 4 compares performance with the full set of 103 variables and with the reduced 20-variable subset. Across all metrics, the impact of dimensionality reduction is modest. LightGBM, for instance, retains essentially the same test AUROC (0.88 with 103 variables, 0.87 with 20) and F1 (0.67 vs. 0.67). XGBoost shows the highest recall in both configurations, an asset when missing positive cases is clinically unacceptable. LR and the MLP improve the balance between precision and recall after reduction, suggesting that the larger feature set introduced some noise or redundancy. GB maintains stable performance but appears to benefit from the broader pool of predictors.

**Table 4. Performance of the five best models with the full set of 103 variables and with the 20 most informative variables. Values are reported as *train* / *test*.**

| Metric | LightGBM | XGBoost | GB | LR | MLP |
|---|---|---|---|---|---|
| Accuracy (103 vars) | 0.87 / 0.84 | 0.84 / 0.81 | 0.86 / 0.85 | 0.86 / 0.85 | 0.85 / 0.85 |
| Accuracy (20 vars) | 0.86 / 0.84 | 0.82 / 0.80 | 0.86 / 0.84 | 0.83 / 0.84 | 0.84 / 0.84 |
| AUC (103 vars) | 0.92 / 0.88 | 0.91 / 0.87 | 0.91 / 0.88 | 0.90 / 0.89 | 0.90 / 0.88 |
| AUC (20 vars) | 0.92 / 0.87 | 0.90 / 0.87 | 0.90 / 0.88 | 0.86 / 0.85 | 0.88 / 0.88 |
| Precision (103 vars) | 0.81 / 0.75 | 0.67 / 0.63 | 0.89 / 0.83 | 0.80 / 0.77 | 0.79 / 0.77 |
| Precision (20 vars) | 0.79 / 0.73 | 0.64 / 0.60 | 0.88 / 0.83 | 0.74 / 0.78 | 0.80 / 0.78 |
| Recall (103 vars) | 0.66 / 0.61 | 0.77 / 0.73 | 0.54 / 0.53 | 0.62 / 0.63 | 0.61 / 0.64 |
| Recall (20 vars) | 0.66 / 0.61 | 0.77 / 0.75 | 0.54 / 0.53 | 0.53 / 0.56 | 0.55 / 0.57 |
| F1-score (103 vars) | 0.73 / 0.67 | 0.72 / 0.67 | 0.67 / 0.65 | 0.70 / 0.69 | 0.69 / 0.70 |
| F1-score (20 vars) | 0.72 / 0.67 | 0.70 / 0.67 | 0.67 / 0.64 | 0.62 / 0.65 | 0.65 / 0.66 |

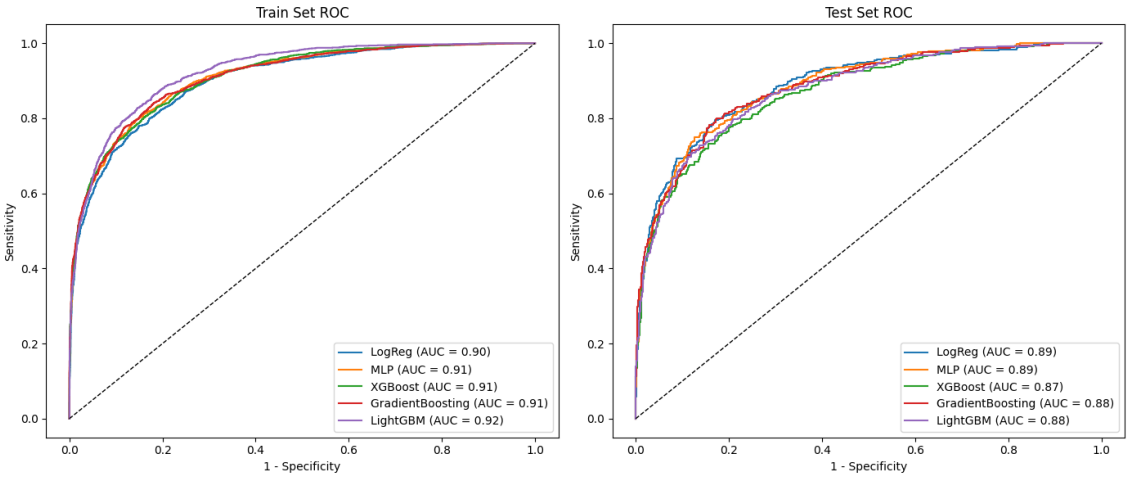


**Figure 7. Receiver-operating-characteristic (ROC) curves for all models without feature selection. Training curves cluster in the upper-left quadrant (AUROC 0.90–0.92), while test curves remain high (0.87–0.89) without signs of overfit; LightGBM retains the largest area under the curve in both splits.**

Balancing these results, LightGBM emerges as the preferred predictor. It combines the highest AUROC and $F_1$-score with stable behaviour between training and testing, and an inherent ability to rank variables for clinical interpretation. XGBoost remains attractive when recall must be prioritised, while LR offers full transparency at the cost of a small performance gap.

Figure 7 shows the ROC curves for the full-feature models. Training curves cluster near the upper-left corner (AUROC 0.90–0.92); the test curves remain high (0.87–0.89) with similar shapes, indicating no obvious over-fitting. LightGBM maintains the largest area under the curve across the entire operating range.

## 5. Conclusions

This study developed a comprehensive machine learning framework for sepsis mortality prediction that addresses critical challenges in clinical implementation while achieving robust predictive performance. Our systematic evaluation of five classification algorithms—LightGBM, XGBoost, Gradient Boosting, Logistic Regression, and Multi-Layer Perceptron—using MIMIC-IV data demonstrates that all optimized models achieved strong discriminative power for 30-day mortality prediction (AUROC $\geq$ 0.87 on independent test sets), with LightGBM emerging as the optimal balance of accuracy (AUROC 0.88), precision-recall trade-off (F1 0.67), and computational efficiency.

First, we demonstrated that feature selection can reduce the predictor set from 103 to 20 clinical variables while preserving predictive performance across all algorithm families, directly addressing the practical constraint of feature availability in real-time clinical settings. Second, our evaluation of algorithm-specific class-imbalance mitigation strategies revealed that tailored approaches for each model type significantly enhance minority class detection capabilities, with ensemble methods (XGBoost, LightGBM) showing superior performance when optimized with appropriate class-weighting schemes. Third, the identification of consistent feature importance patterns across diverse algorithmic approaches provides robust clinical validation of our findings. The consistent emergence of hemodynamic instability markers (heart rate variability, mean arterial pressure) and metabolic dysfunction indicators (lactate levels, pH extremes) as the most predictive variables across all models not only aligns with established sepsis pathophysiology but also offers actionable insights for clinical decision-making that transcend specific algorithmic choices. Fourth, our performance evaluation framework, incorporating discrimination metrics, calibration analysis, and interpretability assessments, establishes reproducible benchmarks that enable comparisons with future research while ensuring clinical relevance.

The clinical implications of these findings are substantial. In a practical workflow, a tool based on this model would not only provide an initial mortality risk score but could also highlight the specific patient variables (e.g., 'rising lactate', 'low blood pressure') driving that risk. This transforms the model from a simple "black box" predictor into an interpretable assistant. It allows clinicians to move from a generic risk score to actionable information, guiding immediate diagnostic or therapeutic decisions and enhancing trust in the system by transparently aligning its reasoning with established medical practice.

Looking toward clinical implementation, the validated 20-variable feature set consists entirely of routinely collected ICU parameters, eliminating barriers related to specialized testing or data unavailability. The framework's modular design allows healthcare institutions to select the most appropriate algorithm based on their specific computational infrastructure and interpretability requirements, while maintaining confidence in predictive performance through our established benchmarks. Future validation in prospective clinical trials and multi-center studies will be essential to confirm generalizability and clinical utility across diverse patient populations and healthcare settings. This work supports the development of sepsis prediction models suitable for clinical use, showing that machine learning can assist in real-world medical decision-making.

## Acknowledgments

Felipe gratefully acknowledges the financial support provided by Nubank S.A. and ITAEx.

## References

Al Omar, S., Alshraideh, J. A., Oweidat, I., Al Qadire, M., Khalaf, A., Sumaqa, Y. A., Al-Mugheed, K., Alabdullah, A. A. S., and Abdelaliem, S. M. F. (2024). Mortality of patients with sepsis in intensive care units at tertiary hospitals in jordan: Prospective cohort study. *Medicine*, 103(43):e40169.

Arbous, S. M., Termorshuizen, F., Brinkman, S., de Lange, D. W., Bosman, R. J., Dekkers, O. M., and de Keizer, N. F. (2024). Three-year mortality of icu survivors with sepsis, an infection or an inflammatory illness: an individually matched cohort study of icu patients in the netherlands from 2007 to 2019. *Critical Care*, 28(1):374.

Bao, C., Deng, F., and Zhao, S. (2023). Machine-learning models for prediction of sepsis patients mortality. *Medicina Intensiva (English Edition)*, 47(6):315–325.

Han, Y., Xie, X., Qiu, J., Tang, Y., Song, Z., Li, W., and Wu, X. (2025). Early prediction of sepsis associated encephalopathy in elderly icu patients using machine learning models: a retrospective study based on the mimic-iv database. *Frontiers in Cellular and Infection Microbiology*, 15:1545979.

Huayanay, A., Bazán, J. L., and Russo, C. M. (2025). Performance of evaluation metrics for classification in imbalanced data. *Computational Statistics*, 40(3):1447–1473.

Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L., and Mark, R. (2024). Mimic-iv (version 3.1). physionet.

Karakike, E., Kyriazopoulou, E., Tsangaris, I., Routsi, C., Vincent, J.-L., and Giamarellos-Bourboulis, E. J. (2019). The early change of sofa score as a prognostic marker of 28-day sepsis mortality: analysis through a derivation and a validation cohort. *Critical care*, 23:1–8.

Meng, C., Trinh, L., Xu, N., Enouen, J., and Liu, Y. (2022). Interpretability and fairness evaluation of deep learning models on mimic-iv dataset. *Scientific Reports*, 12(1):7166.

Shan, W., Sun, D., and Liu, Z.-P. (2024). Predicting sepsis onset in icu patients using machine learning and feature section: A case study of mimic-iv data. In *2024 IEEE International Conference on Medical Artificial Intelligence (MedAI)*, pages 546–551. IEEE.

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., et al. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*, 315(8):801–810.

Yu, Z., Fang, L., and Ding, Y. (2025). Explainable machine learning model for prediction of 28-day all-cause mortality in immunocompromised patients in the intensive care unit: a retrospective cohort study based on mimic-iv database. *European Journal of Medical Research*, 30(1):358.