MACHINE LEARNING AND CLINICAL DATA: AN APPROACH TO MORTALITY PREDICTION IN ICU SEPSIS PATIENTS

JOHAN ALFONSO CASTILLO CABALLERO

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FISICOMECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2025

MACHINE LEARNING AND CLINICAL DATA: AN APPROACH TO MORTALITY PREDICTION IN ICU SEPSIS PATIENTS

JOHAN ALFONSO CASTILLO CABALLERO

Degree work presented as a requirement to qualify for the title of Electronic Engineer

Advisor
Camilo Andres Santos Ortiz
Electronic Engineer

Co-advisor
Carlos Augusto Fajardo Ariza
PhD in Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FISICOMECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2025

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to those who contribute to the collection, development, and management of the MIMIC-IV and eICU databases, whose dedication and effort have provided the scientific community with an invaluable resource.

Gratitude is expressed to the Connectivity and Signal Processing (CPS) research group and Universidad Industrial de Santander for its unwavering institutional support and for providing an inspiring academic environment, elements that have been fundamental to the development of this research.

I am profoundly grateful to my mentors, whose guidance and collaboration have been instrumental in achieving a rigorous and high-quality work, inspiring me to continue exploring new frontiers in science.

Finally, my heartfelt gratitude to my mother and aunts, whose unconditional support, love, and encouragement have been a constant source of strength throughout this journey. Their belief in me, along with the unwavering support of my colleagues, friends, and family, has been essential in completing this stage and continues to inspire me toward new achievements and scientific contributions.

CONTENT

LIST OF FIGURES

page.

LIST OF TABLES

page.

LIST OF APPENDICES

RESUMEN

TÍTULO APRENDIZAJE AUTOMÁTICO Y DATOS CLÍNICOS: UN ENFOQUE EN LA PREDICCIÓN DE LA MORTALIDAD DE PACIENTES CON SEPSIS EN UCI [*]

AUTOR: JOHAN ALFONSO CASTILLO CABALLERO [**]

PALABRAS CLAVE: SEPSIS, UCI, APRENDIZAJE AUTOMÁTICO, MORTALIDAD, MARCAS DE TIEMPO

DESCRIPCIÓN: La sepsis es una condición crítica y potencialmente mortal que se presenta comúnmente en las Unidades de Cuidados Intensivos (UCI). Los profesionales de la salud enfrentan desafíos significativos no solo debido al gran volumen de datos clínicos, sino también por la complejidad inherente de la enfermedad y su naturaleza sistémica. Estos factores crean un entorno desafiante que complica los procesos de toma de decisiones, especialmente cuando se trata de predecir la mortalidad de los pacientes. Dada la importancia crítica de la predicción temprana de la mortalidad para mejorar los resultados de los pacientes, esta investigación tiene como objetivo predecir la mortalidad en pacientes con sepsis en UCI con 12, 24 y 48 horas de anticipación mediante modelos de aprendizaje automático basados en datos clínicos. El estudio se llevó a cabo utilizando la base de datos Medical Information Mart for Intensive Care, que incluye datos de 7,511 pacientes con sepsis en UCI de un solo hospital, y la base de datos Electronic Intensive Care Unit Collaborative Research, que contiene datos de 3,786 pacientes con sepsis en UCI de múltiples hospitales. Se evaluaron ocho modelos de aprendizaje automático supervisado basados en el área bajo la curva, donde Light Gradient Boosted Machine demostró el mejor rendimiento en todos los puntos críticos de tiempo. Además, superó al Sequential Organ Failure Assessment en la predicción de la mortalidad. Esta investigación resalta el potencial del aprendizaje automático para mejorar la predicción de la mortalidad en pacientes con sepsis en UCI, permitiendo una toma de decisiones oportuna y, en última instancia, mejorando los resultados de los pacientes.

---

[*] Trabajo de grado

[**] Facultad de Ingeniería Fisicomecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunucaciones. Ingeniería Electrónica. Director: Camilo Andres Santos Ortiz. Ingeniero Electrónico. Codirector: Carlos Augusto Fajardo Ariza. Doctor en Ingeniería.

ABSTRACT

TITLE: MACHINE LEARNING AND CLINICAL DATA: AN APPROACH TO MORTALITY PREDICTION IN ICU SEPSIS PATIENTS [*]

AUTOR: JOHAN ALFONSO CASTILLO CABALLERO [**]

Keywords: SEPSIS, ICU, MACHINE LEARNING, MORTALITY, TIMESTAMPS.

Description: Sepsis is a critical, life-threatening condition commonly encountered in Intensive Care Units (ICUs). Healthcare professionals face significant challenges not only from the vast volume of clinical data but also due to the condition's inherent complexity and systemic nature. These factors create a challenging environment that complicates decision-making processes, especially when predicting patient mortality. Given the critical importance of early mortality prediction in improving patient outcomes, this research aims to predict mortality for ICU sepsis patients at 12, 24, and 48 hours in advance through machine learning models based on clinical data. The study was conducted with the Medical Information Mart for Intensive Care database, which includes data from 7,511 ICU sepsis patients from a single hospital, and the Electronic Intensive Care Unit Collaborative Research database, which contains data from 3,786 ICU sepsis patients across multiple hospitals. Eight supervised machine learning models were evaluated based on the area under the curve, where Light Gradient Boosted Machine demonstrated the best performance across all critical timestamps. It also outperformed the Sequential Organ Failure Assessment score in predicting mortality. This research underscores the potential of machine learning to advance mortality prediction for ICU sepsis patients, enabling timely decision-making and ultimately improving patient outcomes.

---

[*] BSc Thesis

[**] Facultad de Ingeniería Fisicomecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunucaciones. Ingeniería Electrónica. Advisor: Camilo Andres Santos Ortiz. Electronic Engineer. Co-advisor: Carlos Augusto Fajardo Ariza. PhD in Engineering.

INTRODUCTION

Sepsis is a severe response to infection that can lead to organ failure and death, posing a major challenge in intensive care due to its high mortality rate and complex treatment[1]. In 2017, an alarming 48.9 million cases of sepsis and 11 million deaths attributed to this condition were reported worldwide, significantly surpassing the number of deaths caused by myocardial infarction, lung cancer, breast cancer, and prostate cancer combined[2][3][4]. In Colombia, the prevalence reaches 18.6 % in intensive care units in major cities[5]. This challenge is further compounded by the large volume of clinical data generated in ICU settings, where the high patient volume and the extensive array of variables recorded for each, coupled with the inherent complexity of the condition, present a considerable challenge for healthcare professionals tasked with gathering, interpreting, and leveraging this information efficiently to make informed clinical decisions[6].

---

[1] SINGER, Mervyn et al. *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*. JAMA, 315(8): 801–810, Feb. 2016. ISSN: 0098-7484. DOI: 10.1001/jama.2016.0287.

[2] SIEGEL, Rebecca L. et al. *Cancer Statistics, 2021*. CA: A Cancer Journal for Clinicians, 71(1): 7–33, 2021. DOI: 10.3322/caac.21654.

[3] RUDD, Kristina E. et al. *Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study*. *The Lancet*, 395(10219): 200–211, 2020. DOI: 10.1016/S0140-6736(19)32989-7.

[4] YEH, Robert W. et al. *Population Trends in the Incidence and Outcomes of Acute Myocardial Infarction*. *New England Journal of Medicine*, 362(23): 2155–2165, 2010. DOI: 10.1056/NEJMoa0908610.

[5] RODRÍGUEZ, Ferney et al. *The epidemiology of sepsis in Colombia: A prospective multicenter cohort study in ten university hospitals**. *Critical Care Medicine*, 39(7), 2011. DOI: 10.1097/CCM.0b013e318218a35e.

[6] CELI, Leo A. et al. *"Big dataïn the intensive care unit. Closing the data loop*. *American Journal of Respiratory and Critical Care Medicine*, 187(11): 1157–1160, Jun. 2013. ISBN: 1535-4970; 1073-449X. DOI: 10.1164/rccm.201212-2311ED.

Despite medical advances, early diagnosis and accurate prediction of sepsis progression remain a significant challenge. Current tools, such as the Acute Physiology and Chronic Health Evaluation (APACHEIV)[7], Sequential Organ Failure Assessment (SOFA)[8], quick Sequential Organ Failure Assessment (qSOFA)[9], Simplified Acute Physiology Score (SAPS2)[10], and Systemic Inflammatory Response Syndrome (SIRS2)[11], have contributed to the management of this disease, but they have limitations in their ability to predict adverse outcomes in a timely manner.

Several studies have applied artificial intelligence (AI) approaches to predict mortality in critically ill patients, with different methodologies. Mohamadlou et al. (2019) developed a multicenter model based on Gradient Boosted Trees (GBT), specifically designed to predict mortality in general at 12, 24, and 48 hours before the outcome. While the model successfully leveraged data from multiple hospitals and achieved an average area under the curve (AUROC) of 0.94[12], it lacked a specific focus on sepsis patients in the ICU.

---

7 ZIMMERMAN, Jack E. et al. *Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients*. Critical Care Medicine*, 34(5), 2006. ISBN: 1530-0293. DOI: 10.1097/01.CCM.0000215112.84523.F0.

8 VINCENT, J. et al. *The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure*. *Intensive Care Medicine*, 22(7): 707–710, 1996. ISBN: 1432-1238. DOI: 10.1007/BF01709751.

9 SINGER, Mervyn et al. *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*. JAMA, 315(8): 801–810, Feb. 2016. ISSN: 0098-7484. DOI: 10.1001/jama.2016.0287.

10 LE GALL, Jean-Roger; LEMESHOW, Stanley; SAULNIER, Fabienne. *A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study*. *JAMA*, 270(24): 2957–2963, Dec. 1993. ISSN: 0098-7484. DOI: 10.1001/jama.1993.03510240069035.

11 WYGANT, Dustin B. et al. *Structured Interview of Reported Symptoms-2nd Edition (SIRS-2): Use and Admissibility in Forensic Mental Health Assessment*. *Journal of Personality Assessment*, 104(2): 265–280, 2022. ISBN: 1532-7752; 0022-3891. DOI: 10.1080/00223891.2021.2006673.

12 MOHAMADLOU, Hamid et al. *Multicenter validation of a machine-learning algorithm for 48-h all-cause mortality prediction*. *Health Informatics Journal*, 26(3): 1912–1925, 2020. DOI: 10.1177/1460458219894494. PMID: 31884847.

Additionally, the study did not utilize publicly available databases, which limits its broader applicability and reproducibility. Bao et al. (2023) applied models like LightGBM and XGBoost to large databases such as MIMIC-IV and eICU, focusing on the mortality of sepsis patients in the ICU. Although their model demonstrated an AUROC of 0.96, it only predicted mortality based on the first 24 hours of ICU data, overlooking the critical timestamps of 12, 24, and 48 hours before the outcome[13]. Li and Liu (2024) proposed a composite model that integrates deep neural networks and similarity graphs to predict in-hospital mortality in sepsis patients[14]. While their model innovatively incorporates patient similarity data, it fails to focus on ICU patients specifically, lacks the multicenter data aspect, and does not incorporate temporal prediction windows.

In contrast to these studies, our study aims to predict mortality for ICU sepsis patients at 12, 24, and 48 hours in advance through machine learning models based on clinical data. This research is organized into sections, including the current one, which describes the clinical challenges of sepsis in the ICU, emphasizing high mortality rates and the complexities of managing large clinical datasets. The OBJECTIVES section outlines our general objective along with specific aims. The METHODS section details the design of our retrospective study, describing the use of two clinical datasets and explaining the procedures for data collection, preprocessing, and analysis. In the RESULTS section, we present a comparison of performance metrics, including AUC, precision, recall, and F1 scores evaluated at various timestamps. Finally, the DISCUSSION AND CONCLUSION section discusses the implications of our findings and provides recommendations for future research.

---

[13] BAO, C.; DENG, F.; ZHAO, S. *Machine-learning models for prediction of sepsis patients mortality*. *Medicina Intensiva (English Edition)*, 47(6): 315–325, 2023. ISSN: 2173-5727. DOI: 10.1016/j.medine.2022.06.024.

[14] YONG, Li; LIU, Zhenzhou. *Deep learning-based prediction of in-hospital mortality for sepsis*. *Scientific Reports*, 14(1): 372, 2024. ISBN: 2045-2322. DOI: 10.1038/s41598-023-49890-9.

1. OBJECTIVES

1.1. GENERAL OBJECTIVE

To develop a machine learning model that predicts the probability of mortality in ICU sepsis patients using clinical data.

1.2. SPECIFIC OBJECTIVES

To implement data preprocessing techniques that ensure data quality and integrity for analysis and modeling.

To develop at least three machine learning algorithms for predicting the mortality probability of ICU sepsis patients.

To evaluate the performance of the proposed models using the area under the curve (AUC) metric.

2. METHODS

2.1. DATA SOURCES

A retrospective, nested study design was employed, based on two publicly available clinical databases: the Medical Information Mart for Intensive Care IV (MIMIC-IV v3.0)[15] and the electronic Intensive Care Unit Collaborative Research Database (eICU v2.0)[16]. The MIMIC-IV v3.0, developed by Beth Israel Deaconess Medical Center, is a high-quality open-access clinical resource containing data from 364,627 patients between 2008 and 2019. It includes a wide range of clinical information such as patient demographics, vital signs, laboratory results, medications, and parameters related to continuous renal replacement therapy and mechanical ventilation.

The eICU v2.0, a large multicenter resource, provides data from over 200,000 patients admitted to 348 ICUs across the United States during 2014 and 2015. It includes detailed information on vital signs, laboratory test results, medications, and continuous monitoring parameters during ICU stays. Both datasets are accessible via the PhysioNet platform, with users required to complete the Protecting Human Research Participants course to obtain access. Ethical approval was obtained for the use of these datasets, and necessary human subjects training was completed. Data management and analysis were performed using PostgreSQL[17].

---

[15] JOHNSON, Alistair E. W. et al. *MIMIC-IV, a freely accessible electronic health record dataset*. *Scientific Data*, 10(1): 1, 2023. ISBN: 2052-4463. DOI: 10.1038/s41597-022-01899-x.

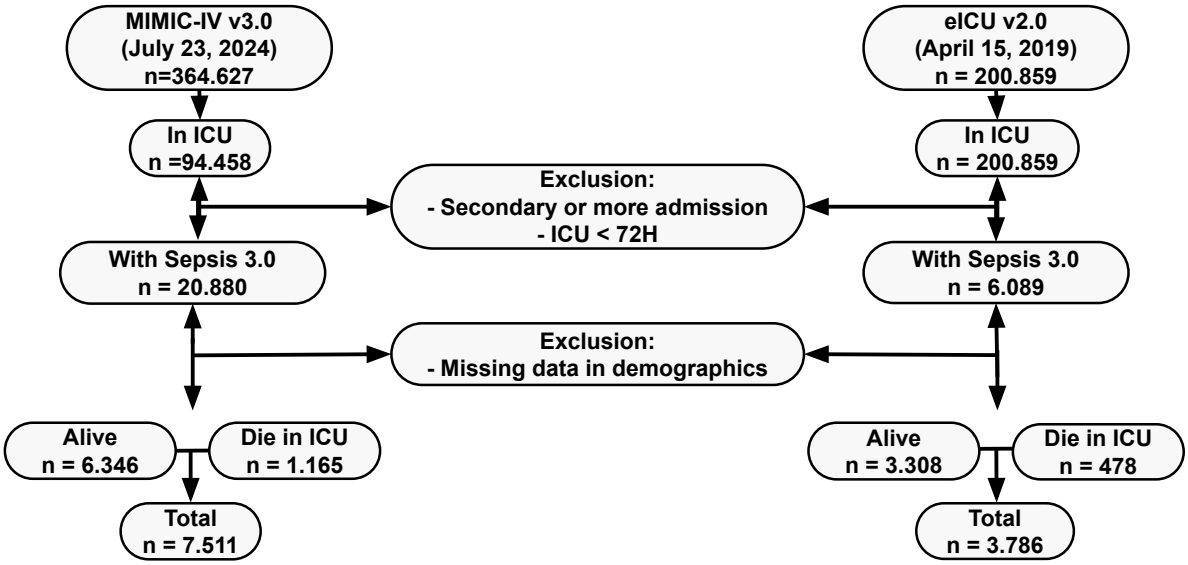
[16] POLLARD, Tom J. et al. *The eICU Collaborative Research Database, a freely available multi-center database for critical care research*. *Scientific Data*, 5(1): 180178, 2018. ISBN: 2052-4463. DOI: 10.1038/sdata.2018.178.

[17] POSTGRESQL GLOBAL DEVELOPMENT GROUP. *PostgreSQL Documentation*. URL: `https://www.postgresql.org/docs/`.

2.2. DATA PREPROCESSING

The data preprocessing involved the application of inclusion and exclusion criteria to ensure the integrity, reliability, and consistency of the dataset. Figure 1 shows a detailed overview of the data extraction process, where strict inclusion criteria were implemented to focus on a clinically relevant population. Only patients diagnosed with sepsis, as defined by the Sepsis-3 criteria[18], were included in the cohort. To minimize potential biases, data were restricted to each patient's first ICU admission, and cases with ICU stays shorter than 72 hours were excluded. Additionally, patient records with incomplete demographic information were removed.

Figure 1. Flow chart of patient selection.



Source: Original work.

---

[18] SINGER, Mervyn et al. *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*. JAMA, 315(8): 801–810, Feb. 2016. ISSN: 0098-7484. DOI: 10.1001/jama.2016.0287.

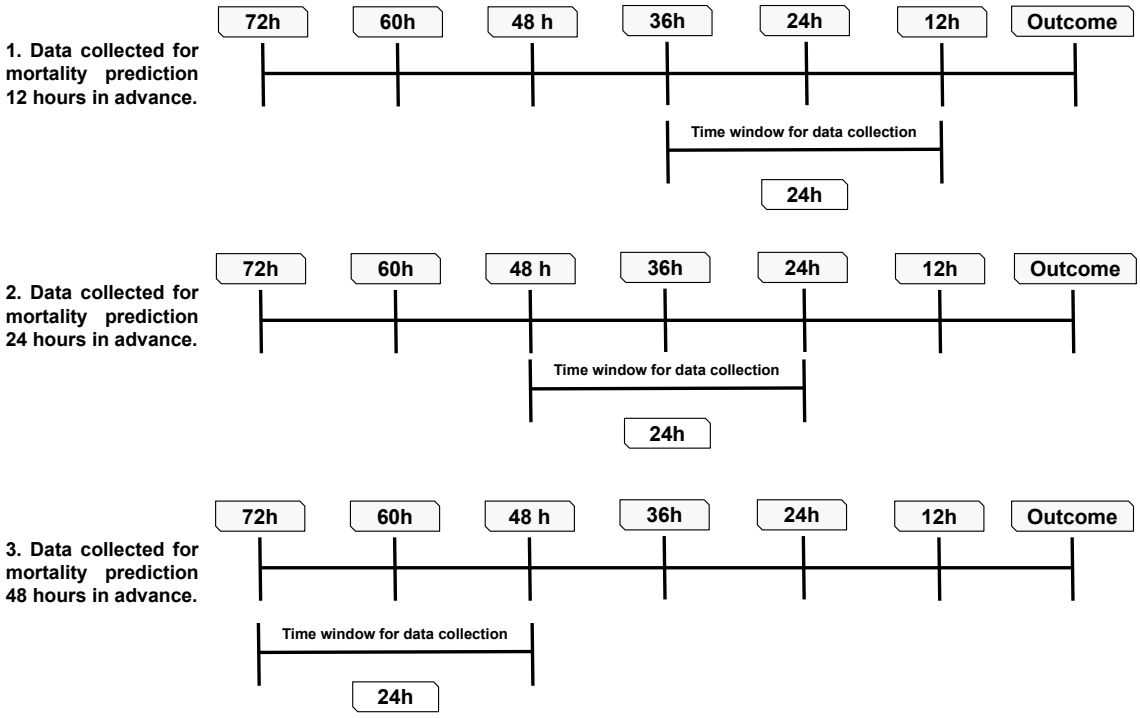
This study cohort comprised 11,297 patients, selected to ensure a robust and clinically relevant analysis. Patient features with more than 20 % missing values were excluded from the cohort's characteristics, while those with 20 % or fewer missing values were imputed with the miceforest library[19], which employs machine learning techniques to reconstruct missing data while preserving the original distribution. From this group, the MIMIC-IV v3.0 dataset provided data from 7,511 patients, representing a single hospital setting and serving as the training cohort. The eICU v2.0 dataset, on the other hand, contributed data from 3,786 patients across multiple hospitals, serving as the testing cohort.

Figure 2 shows the segmentation of the datasets for temporal analysis. For the 12-hour prediction, data from the 36 to 12 hours before the outcome were collected. For the 24-hour prediction, data from the 48 to 24 hours before the outcome were collected. For the 48-hour prediction, data from the 72 to 48 hours before the outcome were collected. This approach, focusing on the 24-hour intervals leading up to each prediction, allowed for a thorough assessment of patient status and provided a solid foundation for making mortality predictions in ICU sepsis patients.

---

[19] WILSON, Sam et al. *Miceforest: Fast, Memory-Efficient Imputation with Random Forests*. Python library for multiple imputation using random forests, version 5.6.0, 2021. `https://github.com/AnotherSamWilson/miceforest`.

Figure 2. Time intervals for data collection in mortality prediction 12, 24, and 48 hours in advance.



Source: Original work.

2.3. MODELS AND METRICS

The features within the time windows for data collection were processed by calculating their maximum, minimum, and mean values, yielding a total of 68 features per dataset, encompassing demographic details, vital signs, laboratory test results, therapeutic interventions, and Glasgow Coma Scale (GCS) scores, which are employed as inputs for model training and testing. A z-score standardization was applied, ensuring that each feature has a mean of zero and a standard deviation of one.

Eight machine learning algorithms were trained and tested in predicting 12, 24, and 48

hours mortality in ICU sepsis patients, with performance evaluated based on the area under the receiver operating characteristic curve (AUC). The models developed included Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), Multi-Layer Perceptron (MLP), Extreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (LGBM). The MIMIC-IV v3.0 dataset was employed for model training, with a stratified 5-fold cross-validation strategy applied. After training, model performance was evaluated on an independent, multicenter dataset from eICU v2.0, which contains data from a wide range of ICUs.

The top-performing model, identified through its AUC score, was selected for further fine-tuning with Optuna [20], an optimization framework aimed at enhancing the model's predictive power. This process ensured that the model's hyperparameters were adjusted to achieve optimal performance. To put the model's capabilities into context, it was compared against the SOFA score across the three critical timestamps: 12, 24, and 48 hours before the outcome.

In addition, to achieve balance between precision and recall, the F1-score was maximized with the MIMIC-IV v3.0 dataset. The optimal thresholds, tailored for each specific timestamp, were subsequently applied to the eICU v2.0 dataset. This allowed for a nuanced comparison of the model's ability to predict mortality with high accuracy at each critical time point. Detailed implementation, including the preprocessing scripts, model training, and code, is provided in Appendix A.

---

[20] AKIBA, Takuya et al. *Optuna: A Next-generation Hyperparameter Optimization Framework*. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631, 2019. ISBN: 9781450362016. Publisher: Association for Computing Machinery, New York, NY, USA. DOI: 10.1145/3292500.3330701.

3. RESULTS

Table 1 shows the performance of eight machine learning algorithms for predicting mortality in ICU sepsis patients. The models were initially trained on a single-center dataset from MIMIC-IV v3.0, employing a stratified 5-fold cross-validation strategy and subsequently evaluated on a diverse, multicenter dataset from eICU v2.0.

Table 1. Comparison of AUC performance of eight models in predicting 12, 24, and 48 hours mortality in ICU sepsis patients from the MIMIC-IV v3.0 and eICU v2.0 datasets. Logistic Regression (LR); Support Vector Machine (SVM); Decision Tree (DT); Random Forest (RF); Gradient Boosting (GB); Multi-Layer Perceptron (MLP); Extreme Gradient Boosting (XGB); and Light Gradient Boosting Machine (LGBM).

Datasets	**Models**	**AUC at different timestamps before the outcome**		
		12h	**24h**	**48h**
MIMIC IV v3.0	**LR**	0.96 (95 % CI: 0.96 - 0.97)	0.94 (95 % CI: 0.93 - 0.94)	0.88 (95 % CI: 0.87 - 0.88)
	SVM	0.96 (95 % CI: 0.96 - 0.97)	0.94 (95 % CI: 0.93 - 0.94)	0.87 (95 % CI: 0.86 - 0.88)
	DT	0.80 (95 % CI: 0.79 - 0.81)	0.75 (95 % CI: 0.74 - 0.76)	0.67 (95 % CI: 0.66 - 0.68)
	RF	0.96 (95 % CI: 0.96 - 0.97)	0.93 (95 % CI: 0.93 - 0.94)	0.87 (95 % CI: 0.86 - 0.88)
	GB	0.97 (95 % CI: 0.96 - 0.97)	0.94 (95 % CI: 0.94 - 0.95)	0.89 (95 % CI: 0.88 - 0.90)
	MLP	0.95 (95 % CI: 0.94 - 0.95)	0.91 (95 % CI: 0.90 - 0.92)	0.84 (95 % CI: 0.83 - 0.85)
	XGB	0.97 (95 % CI: 0.96 - 0.97)	0.94 (95 % CI: 0.94 - 0.95)	0.88 (95 % CI: 0.88 - 0.89)
	LGBM	**0.97 (95 % CI: 0.96 - 0.97)**	**0.94 (95 % CI: 0.94 - 0.95)**	**0.89 (95 % CI: 0.89 - 0.90)**
eICU v2.0	**LR**	0.90 (95 % CI: 0.90 - 0.90)	0.88 (95 % CI: 0.87 - 0.88)	0.82 (95 % CI: 0.81 - 0.82)
	SVM	0.88 (95 % CI: 0.87 - 0.88)	0.84 (95 % CI: 0.83 - 0.84)	0.78 (95 % CI: 0.77 - 0.79)
	DT	0.70 (95 % CI: 0.69 - 0.70)	0.66 (95 % CI: 0.65 - 0.66)	0.60 (95 % CI: 0.59 - 0.61)
	RF	0.90 (95 % CI: 0.89 - 0.90)	0.86 (95 % CI: 0.86 - 0.87)	0.80 (95 % CI: 0.80 - 0.81)
	GB	0.89 (95 % CI: 0.89 - 0.90)	0.87 (95 % CI: 0.86 - 0.87)	0.82 (95 % CI: 0.81 - 0.82)
	MLP	0.87 (95 % CI: 0.87 - 0.88)	0.82 (95 % CI: 0.82 - 0.83)	0.74 (95 % CI: 0.74 - 0.75)
	XGB	0.89 (95 % CI: 0.89 - 0.90)	0.87 (95 % CI: 0.87 - 0.88)	0.81 (95 % CI: 0.80 - 0.81)
	LGBM	**0.90 (95 % CI: 0.90 - 0.91)**	**0.88 (95 % CI: 0.87 - 0.88)**	**0.82 (95 % CI: 0.81 - 0.82)**

Source: Original work.

As expected, the decision tree exhibits lower performance compared to ensemble methods across both the MIMIC-IV v3.0 and eICU v2.0 datasets. In contrast, advanced gradient-boosting models, including GBM, XGB, and LGBM, consistently delivered superior and promising results. Of these, LGBM consistently outperformed the others across all time intervals, achieving the highest AUC values in both the MIMIC-IV and eICU datasets. In the MIMIC-IV dataset, LGBM reached an AUC of 0.97 at 12 hours, 0.94 at 24 hours, and 0.89 at 48 hours. Similarly, in the external eICU dataset, LGBM maintained its lead with AUC values of 0.90 at 12 hours, 0.88 at 24 hours, and 0.82 at 48 hours. These findings underscore the robustness and broad applicability of LGBM, demonstrating its promising potential to enhance mortality prediction for ICU sepsis patients across both datasets and all timestamps.

The results showed that this optimization process maintained LGBM's strong predictive capabilities. To contextualize these findings, LGBM was compared against the SOFA score across various timestamps before the outcome. On the MIMIC-IV single-center dataset, SOFA achieved an AUC of 0.73 at 12 hours before the outcome, substantially lower than LGBM's performance. Similarly, on the eICU multicenter dataset, SOFA's AUC at 12 hours was 0.82, again trailing behind LGBM. Table 2 shows that these differences persisted at 24 and 48 hours, underscoring LGBM's superior predictive performance compared to the SOFA score across all critical timestamps on both single-center and multicenter datasets.

Table 2. Comparison of AUC of model performance LGBM and SOFA in predicting 12, 24, and 48 hours mortality in ICU sepsis patients from the MIMIC IV V3.0 and eICU V2.0 datasets.

Timestamps	Data collection	LGBM (AUC)		SOFA (AUC)	
		MIMIC IV v3.0	eICU v2.0	MIMIC IV v3.0	eICU v2.0
12h	[36h-12h]	0.97 (95 % CI: 0.96 - 0.97)	0.90 (95 % CI: 0.90 - 0.91)	0.73 (95 % CI: 0.72 - 0.74)	0.82 (95 % CI: 0.81 - 0.84)
24h	[48h-24h]	0.94 (95 % CI: 0.93 - 0.94)	0.87 (95 % CI: 0.87 - 0.88)	0.71 (95 % CI: 0.70 - 0.72)	0.80 (95 % CI: 0.78 - 0.81)
48h	[72h-48h]	0.89 (95 % CI: 0.88 - 0.90)	0.82 (95 % CI: 0.81 - 0.82)	0.68 (95 % CI: 0.67 - 0.70)	0.74 (95 % CI: 0.73 - 0.76)

Source: Original work.

To optimize the balance between precision and recall in mortality predictions for ICU sepsis patients, the optimal thresholds for each timestamp before the outcome were determined by maximizing the F1-score, based on the MIMIC-IV v3.0 dataset. Table 3 shows the optimal thresholds, recall, precision, and corresponding F1-scores for the LGBM model in predicting mortality in ICU sepsis patients. Notably, at the 12-hour interval in the eICU v2.0 dataset, the close alignment of precision and recall results in the highest F1-score among the evaluated timestamps. This balanced performance in a diverse, multicenter dataset indicates that the model generalizes across different clinical settings.

Table 3. Comparison of precision, recall and F1-score of the LGBM model in predicting 12, 24 and 48 hours mortality in ICU sepsis patients from the MIMIC IV V3.0 and eICU V2.0 datasets, with threshold variation to maximize the F1-score.

Timestamps	**Data collection**	**Threshold**	**MIMIC IV v3.0**			**eICU v2.0**		
			Precision	**Recall**	**F1 score**	**Precision**	**Recall**	**F1 score**
12h	[36h-12h]	0.21	0.84	0.80	0.82	0.67	0.62	0.64
24h	[48h-24h]	0.34	0.78	0.70	0.74	0.66	0.48	0.56
48h	[72h-48h]	0.22	0.68	0.59	0.63	0.53	0.41	0.46

Source: Original work.

4. DISCUSSION AND CONCLUSION

Validating machine learning models across diverse environments is essential, especially in ICU settings where precise and reliable predictions are critical. The evaluation using the multicenter eICU v2.0 dataset underscores the model's adaptability to varied ICU scenarios, emphasizing the importance of thorough validations. Notably, the LGBM model demonstrated superior performance by improving predictive accuracy by approximately 8 % over current scoring systems, such as SOFA, in predicting mortality in ICU sepsis patients. This approach lays a strong foundation for integrating the model into routine clinical workflows. While challenges such as deployment complexities and potential shifts in performance persist, addressing these hurdles is a crucial step toward advancing critical care through technology, ultimately enhancing decision-making and improving patient outcomes.

The integration of temporal prediction windows at 12, 24, and 48 hours before the outcome represents a significant advancement by aligning predictions with clinical decision-making timelines. While each timestamp covers a 24-hour period, fine-tuning these windows may further enhance model performance. In the eICU v2.0 dataset, the highest AUC was observed at the 12-hour window, with a value of 0.90 and a 95 % confidence interval of 0.90 – 0.91, indicating that the clinical data are more reflective of the immediate pre-outcome state. In contrast, the 48-hour window yielded a slightly lower AUC of 0.82, with a 95 % confidence interval of 0.81 – 0.82. Although this represents a modest reduction in predictive accuracy, it offers the critical benefit of an earlier warning, facilitating more timely clinical interventions. This trade-off underscores the importance of selecting an optimal prediction window in critical care settings. Future research should further investigate these temporal variations to enhance model adaptability and responsiveness, ultimately strengthening its role in managing critical situations.

The current approach, based on single-value metrics, has proven capacity in predicting mortality for ICU sepsis patients. However, integrating waveform data could offer a valuable opportunity to enhance model performance and predictive power. Waveforms provide a more detailed and continuous representation of the patient's physiological state, capturing both temporal and spatial patterns that could further improve predictions. By combining the proven effectiveness of single-value metrics with the depth of waveform data, this integrated approach could pave the way for refining predictions. While challenges remain, such as the need for larger datasets and greater computational resources, this exploration holds the potential to significantly boost the model's performance and enable more precise, timely clinical decision-making. This direction offers an exciting opportunity for future research, with the potential to further improve patient outcomes in the ICU.

BIBLIOGRAPHY

Akiba, Takuya et al.: «Optuna: A Next-generation Hyperparameter Optimization Framework». En: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '19. Anchorage, AK, USA: Association for Computing Machinery, 2019, págs. 2623-2631. ISBN: 9781450362016. DOI: 10.1145/3292500.3330701. https://doi.org/10.1145/3292500.3330701.

Bao, C., F. Deng y S. Zhao: *Machine-learning models for prediction of sepsis patients mortality*. En: *Medicina Intensiva (English Edition)* 47.6 (2023), págs. 315-325. ISSN: 2173-5727. https://doi.org/10.1016/j.medine.2022.06.024.

Celi, Leo A. et al.: *"Big dataïn the intensive care unit. Closing the data loop*. eng. En: *American journal of respiratory and critical care medicine* 187.11 (jun. de 2013). LR: 20220409; GR: 2R01-001659/PHS HHS/United States; JID: 9421642; PMCR: 2014/06/01; 2013/06/04 06:00 [entrez]; 2013/06/04 06:00 [pubmed]; 2013/07/31 06:00 [medline]; 2014/06/01 00:00 [pmc-release]; ppublish, págs. 1157-1160. https://doi.org/10.1164/rccm.201212-2311ED.

Johnson, Alistair E. W. et al.: *MIMIC-IV, a freely accessible electronic health record dataset*. En: *Scientific Data* 10.1 (2023). ID: Johnson2023, pág. 1. ISSN: 2052-4463. https://doi.org/10.1038/s41597-022-01899-x.

Le Gall, Jean-Roger, Stanley Lemeshow y Fabienne Saulnier: *A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study*. En: *JAMA* 270.24 (dic. de 1993), págs. 2957-2963. ISSN: 0098-7484. DOI: 10.1001/jama.1993.03510240069035. https://doi.org/10.1001/jama.1993.03510240069035.

Mohamadlou, Hamid et al.: *Multicenter validation of a machine-learning algorithm for 48-h all-cause mortality prediction*. En: *Health Informatics Journal* 26.3 (2020). PMID: 31884847, págs. 1912-1925. DOI: 10.1177/1460458219894494. eprint: https://doi.org/10.1177/1460458219894494. https://doi.org/10.1177/1460458219894494.

Pollard, Tom J. et al.: *The eICU Collaborative Research Database, a freely available multi-center database for critical care research*. En: *Scientific Data* 5.1 (2018). ID: Pollard2018, pág. 180178. ISSN: 2052-4463. https://doi.org/10.1038/sdata.2018.178.

PostgreSQL Global Development Group: *PostgreSQL Documentation*. https://www.postgresql.org/docs/.

Rodríguez, Ferney et al.: *The epidemiology of sepsis in Colombia: A prospective multicenter cohort study in ten university hospitals**. En: *Critical Care Medicine* 39.7 (2011). ID: 00003246-201107000-00011. https://doi.org/10.1097/CCM.0b013e318218a35e.

Rudd, Kristina E. et al.: *Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study*. En: *The Lancet* 395.10219 (2020). doi: 10.1016/S0140-6736(19)32989-7; 05, págs. 200-211. ISSN: 0140-6736. https://doi.org/10.1016/S0140-6736(19)32989-7.

Siegel, Rebecca L. et al.: *Cancer Statistics, 2021*. En: *CA: A Cancer Journal for Clinicians* 71.1 (2021), págs. 7-33. DOI: https://doi.org/10.3322/caac.21654. eprint: https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21654. https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21654.

Singer, Mervyn et al.: *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*. En: *JAMA* 315.8 (feb. de 2016), págs. 801-810. ISSN: 0098-7484. DOI: 10.1001/jama.2016.0287. eprint: https://jamanetwork.com/journals/jama/articlepdf/2492881/jsc160002.pdf. https://doi.org/10.1001/jama.2016.0287.

Teasdale, Graham y Bryan Jennett: *ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS: A Practical Scale*. En: *The Lancet* 304.7872 (1974). Originally published as Volume 2, Issue 7872, págs. 81-84. ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(74)91639-0. https://www.sciencedirect.com/science/article/pii/S0140673674916390.

Vincent, J. -. et al.: *The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure*. En: *Intensive care medicine* 22.7 (1996). ID: Vincent1996, págs. 707-710. ISSN: 1432-1238. https://doi.org/10.1007/BF01709751.

Wilson, Sam y Contributors: *Miceforest: Fast, Memory-Efficient Imputation with Random Forests*. Ver. 5.6.0. Python library for multiple imputation using random forests. 2021. https://github.com/AnotherSamWilson/miceforest.

Wygant, Dustin B. et al.: *Structured Interview of Reported Symptoms-2nd Edition (SIRS-2): Use and Admissibility in Forensic Mental Health Assessment*. eng. En: *Journal of personality assessment* 104.2 (2022). LR: 20220502; JID: 1260201; 2021/12/07 06:00 [pubmed]; 2022/05/03 06:00 [medline]; 2021/12/06 17:17 [entrez]; ppublish, págs. 265-280. ISSN: 1532-7752, 0022-3891. https://doi.org/10.1080/00223891.2021.2006673.

Yeh, Robert W. et al.: *Population Trends in the Incidence and Outcomes of Acute Myocardial Infarction*. En: *New England Journal of Medicine* 362.23 (2010), págs. 2155-2165. DOI: 10.1056/NEJMoa0908610. eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa0908610. https://www.nejm.org/doi/full/10.1056/NEJMoa0908610.

Yong, Li y Liu Zhenzhou: *Deep learning-based prediction of in-hospital mortality for sepsis*. En: *Scientific Reports* 14.1 (2024). ID: Yong2024, pág. 372. ISSN: 2045-2322. https://doi.org/10.1038/s41598-023-49890-9.

Zimmerman, Jack E. et al.: *Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients**. En: *Critical Care Medicine* 34.5 (2006). ID: 00003246-200605000-00001. ISSN: 1530-0293. https://doi.org/10.1097/01.CCM.0000215112.84523.F0.

APPENDICES

Appendix A. GitHub repository

The source code and supplementary materials related to this project are available in the following GitHub repository:

https://github.com/johancastilloc/Machine-learning-and-clinical-data-an-approach-to-mortality-prediction-in-ICU-sepsis-patients