

PREDICCIÓN DE MORTALIDAD EN TERAPIA INTENSIVA MEDIANTE ALGORITMOS DE MACHINE LEARNING: COMPARACIÓN CON MODELOS TRADICIONALES Y ANÁLISIS DE DESEMPEÑO

MORTALITY PREDICTION IN INTENSIVE CARE USING MACHINE LEARNING ALGORITHMS: COMPARISON WITH TRADITIONAL MODELS AND PERFORMANCE ANALYSIS

AUTORES: BANEGAS, ALEXIS ****; GENOVESE, GABRIELA***; ANTIK, ARIEL **; CORONADO, PAUL*

LUGAR DE TRABAJO: UNIDAD DE TERAPIA INTENSIVA ADULTOS. HOSPITAL GENERAL DE AGUDOS CARLOS G. DURAND -AV. DIAZ VELEZ 5044-C.P. 1405 - CABA, ARGENTINA.

* JEFE DE UNIDAD
** JEFE DE UNIDAD Y COORDINADOR DE RESIDENTES
*** MÉDICA DE PLANTA
**** JEFE DE RESIDENTES



ABSTRACT:

INTRODUCTION: Predicting mortality in intensive care is crucial for decision-making. This study evaluates machine learning algorithms as a complementary tool to traditional scoring systems, analyzing their accuracy in critically ill patients during the first 24 hours of admission.

OBJETIVE: To model machine learning algorithms for predicting mortality in critically ill patients.

MATERIALS AND METHODS: Retrospective observational study using the SATI-Q database. Adult patients with complete data within the first 24 hours of admission were included. Supervised classification algorithms were trained and validated: logistic regression (LR), k-nearest neighbors (KNN), random forest (RF), and artificial neural networks (ANN). Mortality was assessed as the outcome variable, and the following were included as predictor variables: age, SAPS II score, arterial pH, and mechanical ventilation (MV). The sample was divided into training (70%) and testing (30%) groups with cross-validation. Metrics: Accuracy, Precision, Recall, F1 score, and area under the ROC curve (AUC).

RESULTS: 617 patients were included, with a mortality rate of 20%. ANN showed the best discriminative performance (AUC 0.85), followed by RF (0.67), KNN (0.61), and LR (0.60). Accuracy was similar between ANN and LR (0.82). The most relevant predictors in LR were SAPS II and MV.

CONCLUSION: The artificial neural network model demonstrated superior predictive capacity compared to the other algorithms evaluated. Artificial intelligence is being progressively incorporated into medical decision-making processes. However, the use of "black box" algorithms raises a significant ethical dilemma in the field of medicine.

RESUMEN:

INTRODUCCIÓN: La predicción de mortalidad en terapia intensiva es clave para la toma de decisiones. Este estudio evalúa algoritmos de machine learning como herramienta complementaria a los scores tradicionales, analizando su precisión en pacientes críticos durante las primeras 24 horas de ingreso.

OBJETIVO: Modelar algoritmos de machine learning para predecir la mortalidad del paciente crítico.

MATERIALES Y MÉTODOS: Estudio observacional retrospectivo utilizando la base de datos SATI-Q. Se incluyeron pacientes adultos con datos completos dentro de las primeras 24 horas de ingreso. Se entrenaron y validaron algoritmos supervisados de clasificación: regresión logística (RL), k-nearest neighbors (KNN), random forest (RF) y redes neuronales artificiales (ANN). Como variable de resultado se evaluó la mortalidad, y como variables predictoras se incluyeron: edad, SAPS II, pH arterial y asistencia respiratoria mecánica (ARM). Se dividió la muestra en entrenamiento (70%) y testeo (30%) con validación cruzada. Métricas: Accuracy, Precision, Recall, F1-score y área bajo la curva ROC (AUC).

RESULTADOS: Se incluyeron 617 pacientes, con mortalidad del 20%. La ANN mostró el mejor desempeño discriminativo (AUC 0.85), seguida de RF (0.67), KNN (0.61) y RL (0.60). La accuracy fue similar entre ANN y RL (0.82). Los predictores más relevantes en RL fueron SAPS II y ARM.

CONCLUSIÓN: El modelo de red neuronal artificial demostró una capacidad predictiva superior en comparación con los demás algoritmos evaluados. La inteligencia artificial está siendo progresivamente incorporada en los procesos de toma de decisiones médicas. Sin embargo, el uso de algoritmos de tipo "caja negra" plantea un dilema ético significativo en el ámbito de la medicina.

Keywords: Mortality, intensive care, algorithm

Palabras clave: Mortalidad, terapia intensiva, algoritmo

INTRODUCCIÓN

La predicción de mortalidad en pacientes críticos constituye un componente esencial en la práctica de la terapia intensiva, ya que influye directamente en la toma de decisiones clínicas, la planificación terapéutica y la asignación de recursos. Desde

hace décadas, se han desarrollado múltiples sistemas de puntuación, como APACHE II y SAPS II, que permiten estimar el riesgo de muerte a partir de variables clínicas y fisiológicas obtenidas durante las primeras horas de internación. Sin embargo, estos modelos presentan limitaciones importantes. En primer lugar, su capacidad predictiva puede verse afectada por la necesidad de recalibración periódica, especialmente cuando se aplican en poblaciones distintas a aquellas en las que fueron desarrolladas. Además, su estructura basada en modelos estadísticos lineales limita la capacidad para capturar relaciones complejas y no lineales entre variables, lo cual es particularmente relevante en el paciente crítico, donde la interacción entre múltiples sistemas fisiológicos es altamente dinámica.

En los últimos años, el desarrollo de técnicas de machine learning (ML) ha abierto nuevas perspectivas en el análisis de datos clínicos. Estos algoritmos permiten procesar grandes volúmenes de información y detectar patrones complejos, lo que podría traducirse en una mejora en la capacidad de predicción de eventos clínicos relevantes, incluyendo la mortalidad.

En el contexto de la medicina de precisión y la digitalización de los sistemas de salud, la incorporación de modelos basados en inteligencia artificial en unidades de terapia intensiva representa una oportunidad para optimizar la toma de decisiones clínicas y mejorar la estratificación del riesgo desde etapas tempranas de la internación.

El presente estudio tiene como objetivo evaluar el desempeño de diferentes algoritmos de machine learning en la predicción de mortalidad en pacientes críticos, y compararlos con modelos tradicionales, utilizando variables disponibles durante las primeras 24 horas de ingreso.

OBJETIVO

Modelar y evaluar distintos algoritmos de machine learning para la predicción de mortalidad en pacientes críticos internados en unidades de terapia intensiva, comparando su desempeño mediante métricas estándar.

MATERIALES Y MÉTODOS

Se realizó un estudio observacional, retrospectivo, utilizando datos provenientes de la base SATI-Q, que recopila información de pacientes internados en unidades de terapia intensiva.

Se incluyeron pacientes adultos (≥ 18 años) con datos completos registrados dentro de las primeras 24 horas de ingreso a la unidad de terapia intensiva. Se excluyeron aquellos pacientes con datos incompletos en las variables seleccionadas para el análisis.

Variable dependiente: Mortalidad en unidad de terapia intensiva

Variables independientes: Edad - SAPS II - pH arterial - Requerimiento de asistencia respiratoria mecánica (ARM).

Se entrenaron y evaluaron los siguientes algoritmos de clasificación supervisada:

Regresión logística (RL) - K-nearest neighbors (KNN) - Random forest (RF) - Red neuronal artificial (ANN).

Los datos fueron divididos en un conjunto de entrenamiento (70%) y un conjunto de testeo (30%). Las variables continuas fueron normalizadas previo al entrenamiento de los modelos. Se realizó validación cruzada y optimización de hiperparámetros para mejorar el rendimiento de los algoritmos. El desempeño de los modelos se evaluó mediante: Accuracy - Precision - Recall - F1 score - Área bajo la curva ROC (AUC)

El análisis se realizó utilizando el lenguaje de programación Python, con las bibliotecas Pandas, Scikit-learn y Matplotlib.

Consideraciones éticas: El estudio se llevó a cabo utilizando una base de datos anonimizada, sin intervención sobre los pacientes, garantizando la confidencialidad de la información de acuerdo con las normativas vigentes.

RESULTADOS

Se incluyeron un total de 617 pacientes. La edad mediana fue de 58 años (RIC 40-69), con un APACHE II de 14 (RIC 9-20). El 43% de los pacientes requirió asistencia respiratoria mecánica y la mortalidad global fue del 20%.

En el conjunto de testeo, los modelos mostraron el siguiente rendimiento:

Regresión logística: Accuracy 0.82 - AUC 0.60

KNN: Accuracy 0.80 - AUC 0.61

Random forest: Accuracy 0.73 - AUC 0.67

Red neuronal artificial: Accuracy 0.82 - AUC 0.85 (Fig. 1).

La red neuronal artificial presentó el mejor desempeño global, destacándose especialmente en la capacidad de discriminación, evidenciada por un AUC significativamente superior al resto de los modelos. (Fig. 2).

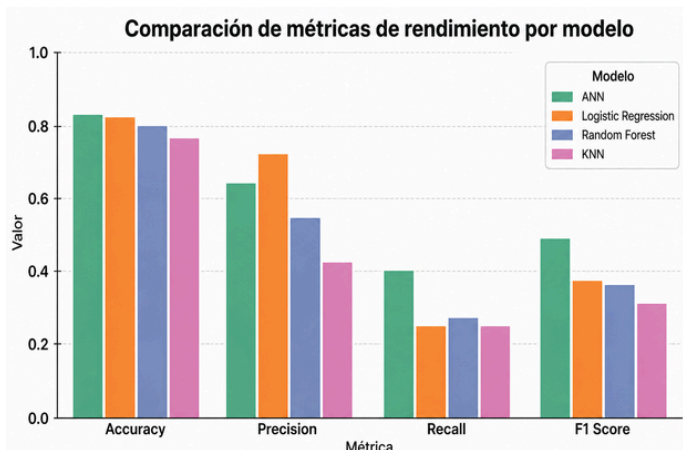


Fig. 1. Comparación de las métricas de rendimiento (Accuracy, AUC, Precision, Recall y F1-score) entre los modelos de machine learning evaluados.

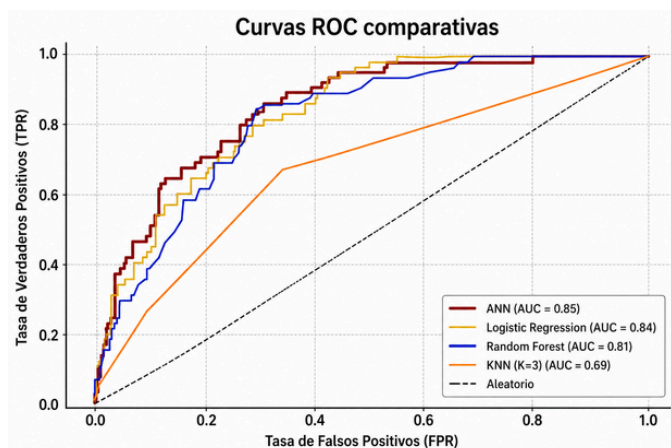


Fig. 2. Curvas ROC de los modelos utilizados para la predicción de mortalidad en pacientes críticos.

En el análisis de la regresión logística, las variables con mayor impacto en la predicción de mortalidad fueron el SAPS II (coeficiente 3.56) y el requerimiento de ARM (coeficiente 0.123), lo que resalta la importancia de la gravedad inicial y el compromiso respiratorio en el pronóstico del paciente crítico.

DISCUSIÓN

Los resultados de este estudio demuestran que los algoritmos de machine learning, en particular las redes neuronales artificiales, pueden mejorar la capacidad predictiva de mortalidad en pacientes críticos en comparación con modelos tradicionales. Estos hallazgos son consistentes con estudios previos que destacan la superioridad de modelos no lineales en contextos clínicos complejos. La superioridad observada en la ANN puede explicarse por su capacidad para modelar relaciones no lineales e interacciones complejas

entre variables clínicas. En el contexto de la terapia intensiva, donde múltiples sistemas fisiológicos interactúan de manera dinámica, esta característica resulta especialmente relevante. Por otro lado, la regresión logística, si bien mostró un buen desempeño en términos de accuracy, presentó una menor capacidad discriminativa, lo que pone en evidencia las limitaciones de los modelos lineales frente a la complejidad del paciente crítico.

No obstante, los modelos tradicionales conservan una ventaja importante en términos de interpretabilidad. A diferencia de las redes neuronales, cuya lógica interna puede resultar opaca, la regresión logística permite identificar el peso relativo de cada variable, facilitando su aplicación clínica.

Este aspecto introduce un desafío central en la implementación del machine learning en medicina: el equilibrio entre precisión predictiva y explicabilidad. Los modelos de tipo “caja negra” plantean interrogantes éticos y legales, particularmente cuando sus resultados influyen en decisiones clínicas críticas. En este sentido, el objetivo no es reemplazar al médico, sino complementar el proceso de toma de decisiones mediante herramientas basadas en datos.

Entre las limitaciones del estudio se encuentran su diseño retrospectivo, el número limitado de variables incluidas y la ausencia de validación externa, lo que podría afectar la generalización de los resultados.

Futuros estudios deberían incorporar variables dinámicas, mayor volumen de datos y validación multicéntrica, con el fin de mejorar la robustez y aplicabilidad de los modelos.

CONCLUSIÓN

Los algoritmos de machine learning, particularmente las redes neuronales artificiales, demostraron una mayor capacidad para predecir la mortalidad en pacientes críticos en comparación con otros modelos evaluados. Estos hallazgos respaldan su potencial como herramientas complementarias a los sistemas de puntuación tradicionales. No obstante, su implementación en la práctica clínica requiere validación externa, adecuada interpretación y consideración de aspectos éticos. La integración de inteligencia artificial en terapia intensiva representa un paso hacia una medicina más precisa, predictiva y personalizada.

Conflictos de Interés :Los autores del artículo declaran que no hay ningún conflicto de interés al publicar el manuscrito.

Recibido: 28/03/2026 Aceptado: 15/04/2026

BIBLIOGRAFIA

1. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
 2. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis. *Intensive Care Med*. 2020;46:383-400.
 3. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035.
 4. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-829.
 5. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II). *JAMA*. 1993;270(24):2957-2963.
 6. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347-1358.
 7. Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care*. 2019;23:284.
 8. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56.
-