

ARTÍCULO ORIGINAL

Sistema inteligente para la predicción de mortalidad en Unidades de Cuidados Intensivos

Intelligent system for mortality prediction in Intensive Care Units

María del Rocío Vicente González
mvgonzalez@uclv.cu • <https://orcid.org/0009-0007-8841-0950>

María Matilde García Lorenzo
mmgarcia@uclv.edu.cu • <https://orcid.org/0000-0002-1663-5794>

Rafael Fernández Fleites
rffleites@uclv.cu • <https://orcid.org/0009-0009-5947-2567>

Roberto Vicente Rodríguez
rvicenter@uclv.cu • <https://orcid.org/0009-0002-1414-8482>

UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS

Recibido: 2025-09-04 • Aceptado: 2025-10-18

RESUMEN

La predicción temprana de mortalidad en Unidades de Cuidados Intensivos (UCI) mediante técnicas de aprendizaje automático se ha convertido en una herramienta clave para apoyar la toma de decisiones clínicas en entornos altamente dinámicos y críticos. En este trabajo se desarrolla un modelo de red neuronal Perceptrón Multicapa (MLP) para predecir la mortalidad en la UCI del Hospital Universitario "Arnaldo Milián Castro" de Villa Clara, Cuba. Se aplicaron técnicas de preprocesamiento para manejar inconsistencias en los datos clínicos, así como métodos de balanceo como BorderlineSMOTE para abordar el desbalance de clases. El modelo alcanzó métricas destacables en pruebas de campo: sensibilidad superior al 85%, área bajo la curva ROC (AUC-ROC) de 0,93 y precisión global elevada. Además, se implementó el método LIME (Local Interpretable Model-Agnostic Explanations) para mejorar la interpretabilidad de las predicciones. Los resultados evidencian el potencial del sistema para identificar pacientes en riesgo con un mínimo margen de error, constituyendo una herramienta de apoyo valiosa para el personal médico en entornos de cuidados intensivos.

Palabras clave: Predicción de mortalidad; Unidad de Cuidados Intensivos; Perceptrón Multicapa; Inteligencia artificial explicable; LIME.

ABSTRACT

Early mortality prediction in Intensive Care Units (ICU) using machine learning techniques has become a key tool for supporting clinical decision-making in highly dynamic and critical healthcare environments. In this study, a Multilayer Perceptron (MLP) neural network model was developed to predict mortality in the ICU of the "Arnaldo Milián Castro" University Hospital in Villa Clara, Cuba. Preprocessing techniques were applied to handle inconsistencies in clinical data, along with balancing methods such as BorderlineSMOTE to address class imbalance. The model achieved notable metrics in field tests: sensitivity above 85%, area under the ROC curve (AUC-ROC) of 0.93, and high overall accuracy. Additionally, the LIME method (Local Interpretable Model-Agnostic Explanations) was implemented to enhance the interpretability of predictions. The results demonstrate the system's potential to identify at-risk patients with minimal error margin, providing a valuable support tool for medical staff in intensive care settings.

Keywords: Mortality prediction; Intensive Care Unit; Multilayer Perceptron; Explainable artificial intelligence; LIME.

INTRODUCCIÓN

La predicción temprana de mortalidad en pacientes ingresados en Unidades de Cuidados Intensivos (UCI) representa un desafío crítico en la medicina intensiva. Los sistemas tradicionales de puntuación, como APACHE II y SOFA, presentan limitaciones significativas debido a su enfoque lineal y falta de adaptabilidad a poblaciones específicas o entornos con datos incompletos (Nistal-Nuño, 2021). En contraste, los sistemas basados en inteligencia artificial (IA) emergen como una solución prometedora, capaces de procesar grandes volúmenes de datos heterogéneos para generar predicciones precisas y personalizadas (Belciug & Gorunescu, 2020).

En el contexto cubano, el Hospital Universitario "Arnaldo Milián Castro" enfrenta desafíos particulares, como un volumen limitado de datos, falta de estandarización en variables clave y un marcado desbalance entre clases (supervivencia vs. mortalidad). Estos factores dificultan la implementación de modelos predictivos robustos (Johnson et al., 2016).

El presente trabajo tiene como objetivo desarrollar un sistema predictivo de mortalidad en UCI basado en IA, específicamente mediante el uso de redes neuronales Perceptrón Multicapa (MLP). Se aborda el preprocesamiento de datos clínicos, la construcción y optimización del modelo, y la evaluación de su rendimiento en condiciones reales, complementado con técnicas de explicabilidad para facilitar su adopción clínica (Shneiderman, 2020).

METODOLOGÍA

Diseño de la investigación

Se realizó un estudio observacional retrospectivo basado en datos clínicos de pacientes ingresados en la UCI del Hospital Universitario "Arnaldo Milián Castro" entre 2017 y 2024. La investigación siguió un enfoque cuantitativo, que consta de cinco etapas mostradas en la Figura1.

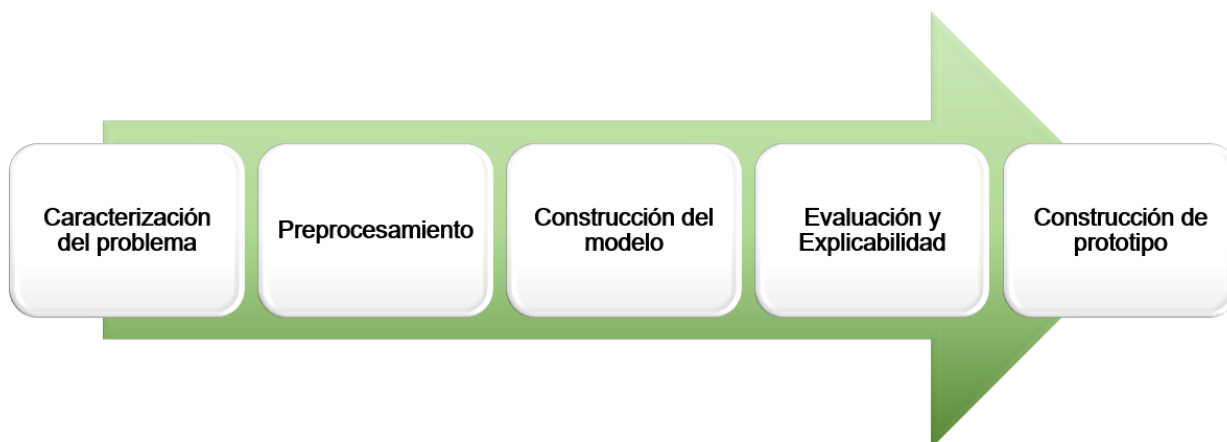


Figura 1. Metodología de la Investigación

Caracterización del Problema

La base inicial contaba con 3.398 registros y 34 atributos (10 numéricos y 24 categóricos), uno de los cuales es la clase para pronosticar "Murió". Se cuenta con 2 718 pacientes en UCI que no mueren y 680 que mueren.

La base de casos analizada presenta diversos problemas que afectan la calidad y confiabilidad de los datos para el desarrollo de modelos predictivos. En primer lugar, se observa un marcado desbalance entre clases, con una proporción significativamente menor de pacientes fallecidos respecto a los que sobreviven. Además, se identificaron múltiples inconsistencias en los registros, incluyendo valores faltantes o nulos en variables clínicas clave, registros con valor cero en atributos que deberían tener mediciones válidas, y discrepancias entre variables relacionadas. También se detectaron problemas de formato, como campos con cadenas de texto no estandarizadas, valores duplicados con variaciones por espacios o caracteres invisibles, y registros en lenguaje natural sin codificación uniforme. Asimismo, existen incoherencias temporales, como fechas de egreso anteriores a las de ingreso o tiempos de estancia negativos, lo que compromete la integridad de la información temporal. Estos desafíos subrayan la necesidad de un preprocesamiento riguroso antes de emplear los datos en modelos de aprendizaje automático.

Preprocesamiento

Reducción de dimensionalidad de la base de casos a partir de la opinión de expertos

Tras un análisis de expertos y eliminación de variables irrelevantes, se conservaron ocho atributos de los 34 iniciales: edad, sexo, raza, APACHE, Categoría/Diagnóstico/Especificación, Tiempo exacto de ventilación, Tiempo de estancia, Murió (clase del estudio).

Solución de inconsistencias

Una de las primeras acciones consistió en abordar los registros de edad con valor igual a cero. En lugar de eliminar estos casos y reducir el tamaño de la muestra, se optó por una estrategia de imputación basada en la mediana de las edades registradas en la población total. La mediana, a diferencia de la media, no se ve influenciada por valores extremos o atípicos, lo que la convierte en una medida de tendencia central más robusta para variables como la edad en contextos clínicos.

Uno de los desafíos más complejos fue la imputación de los valores del índice APACHE II registrados como cero, un problema que surgía de limitaciones logísticas en la disponibilidad de equipos o reactivos para su cálculo en el momento del ingreso. Dada la importancia pronóstica de este score en la evaluación de la gravedad del paciente en la UCI, no era viable simplemente ignorar estos registros. Se diseñó una solución basada en aprendizaje automático, utilizando un algoritmo de Random Forest (Azar et al., 2014) entrenado con los casos que sí contaban con un valor APACHE válido.

Los problemas relacionados con el tiempo de estancia en la UCI requirieron un tratamiento particular. Se identificaron casos con duraciones registradas como cero o incluso con valores negativos, incongruencias temporales que invalidaban cualquier análisis basado en esta variable. Para resolver esto, se estableció una jerarquía de sustitución: cuando el tiempo de estancia era inválido, se utilizó el tiempo exacto de ventilación asistida como proxy, dado que este periodo generalmente constituye una parte sustancial de la hospitalización en cuidados intensivos. En los casos donde ambos valores eran cero o inconsistentes, se recurrió nuevamente a la mediana de los tiempos de estancia válidos, asegurando así la coherencia temporal de todos los registros.

Finalmente, se abordó la complejidad de la información diagnóstica, originalmente contenida en un único campo de texto con múltiples tuplas de "Categoría/Diagnóstico/Especificación" concatenadas. Para resolver esta limitación, se desarrolló un proceso de normalización y desagregación que separó cada tupla diagnóstica en atributos individuales: diag1, diag2, diag3, diag4, diag5, diag6 y diag7. Esta transformación no solo facilitó la codificación posterior de los diagnósticos, sino que permitió capturar la multimorbilidad de los pacientes y analizar la influencia de diagnósticos primarios y secundarios en el riesgo de mortalidad.

Reducción de Dimensionalidad mediante Selección de Atributos

Una vez resueltas las inconsistencias básicas en los datos, se procedió a abordar el desafío de la alta dimensionalidad mediante la aplicación sistemática de técnicas de selección de atributos para identificar las variables más informativas y eliminar aquellas redundantes o poco relevantes.

Se aplicaron tres algoritmos complementarios de selección de atributos: Information Gain Ranking Filter, Gain Ratio Feature Evaluator, y ReliefF Ranking Filter (Sánchez Martín, 2024).

Los algoritmos de selección fueron aplicados sobre dos representaciones diferentes de los datos diagnósticos. La primera conservaba la estructura completa de las tuplas "Categoría/Diagnóstico/Especificación", capturando así el máximo detalle disponible. La segunda simplificaba esta información considerando únicamente el campo "Diagnóstico" principal, ofreciendo una representación más general pero potencialmente más estable y generalizable.

El análisis de los resultados de selección de atributos se realizó mediante dos criterios de inclusión progresivos. Primero, se consideraron aquellos atributos que al menos dos de los tres algoritmos coincidían en identificar como

relevantes, con un umbral de importancia superior a 0.01. Este criterio más inclusivo dio lugar a dos conjuntos de variables: para la base con tuplas diagnósticas completas se seleccionaron diag1, diag2, Apache, Edad, Tiempo exacto de ventilación y stay; mientras que para la base con diagnósticos simplificados se mantuvo exactamente el mismo conjunto.

Posteriormente, se aplicó un criterio más estricto, seleccionando únicamente aquellos atributos que los tres algoritmos coincidían en identificar como relevantes. Esta selección conservadora generó conjuntos más reducidos: diag1, Apache, Edad y Tiempo exacto de ventilación para la representación diagnóstica completa, y exactamente las mismas cuatro variables para la representación simplificada.

De esta etapa de preprocesamiento se obtuvieron las bases de casos mostradas en la tabla1.

Tabla1. Bases de casos obtenidas del preprocesamiento

	Tipo de Diagnóstico	Criterio de Selección	Atributos
Categoría-diagnóstico-especificación 1	Tuplas Completas	3 algoritmos	diag1, diag2, Apache, Edad, Tiempo exacto de ventilación, stay
Categoría-diagnóstico-especificación 2	Tuplas Completas	2 algoritmos	diag1, Apache, Edad, Tiempo exacto de ventilación
Diagnóstico 1	Solo Diagnósticos	3 algoritmos	diag1, diag2, Edad, Apache, Tiempo exacto de ventilación, stay
Diagnóstico 2	Solo Diagnósticos	2 algoritmos	diag1, Edad, Apache, Tiempo exacto de ventilación

Balanceo de Clases

Debido al desbalance entre clases (680 fallecidos vs. 2.718 supervivientes), se evaluaron técnicas de sobremuestreo como SMOTE, BorderlineSMOTE y Random Oversampling (Vieira & Rodrigues, 2024). BorderlineSMOTE mostró el mejor rendimiento en la mitigación del sesgo hacia la clase mayoritaria.

Construcción del Modelo

Se compararon los algoritmos de clasificación: K-Nearest Neighbors (KNN), Regresión Logística y MLP para las 4 bases de casos obtenidas del preprocesamiento. El algoritmo MLP aplicado a la base de casos Diagnóstico1 demostró superioridad en sensibilidad y AUC-ROC (Harnecker Fontecilla, 2017) que son los campos que se busca optimizar en el estudio, por lo que MLP queda seleccionado como algoritmo de clasificación base para la construcción del modelo.

La priorización del recall como métrica crítica se fundamenta en la naturaleza del dominio clínico abordado. En el contexto de predicción de mortalidad en Unidades de Cuidados Intensivos, los falsos negativos (pacientes que fallecen pero que el modelo clasifica como sobrevivientes) tienen consecuencias muy graves (Nistal-Nuño, 2021). Un modelo con bajo recall dejaría sin detectar a pacientes en riesgo crítico, comprometiendo intervenciones oportunas. Esta decisión refleja el principio de precaución característico de los entornos de cuidados intensivos, donde es preferible generar alertas adicionales que omitir casos de riesgo real (Shneiderman, 2020).

La selección del AUC-ROC como métrica complementaria responde a las características específicas del conjunto de datos disponible. Dado que la base de casos presenta un marcado desbalance de clases (680 fallecidos frente a 2.718 supervivientes), se requiere una métrica que evalúe la capacidad discriminativa del modelo independientemente del umbral de clasificación (Vieira & Rodrigues, 2024). El AUC-ROC proporciona esta evaluación integral, siendo robusto frente al desbalance de clases, a diferencia de métricas como la precisión global que pueden verse artificialmente infladas por la clase mayoritaria (Fernández et al., 2018).

Tras la elección del algoritmo de clasificación se optimizaron los hiperparámetros mediante el algoritmo RandomizedSearchCV (Yang & Shami, 2020), definiendo un espacio de búsqueda que incluyó arquitecturas de red, funciones de activación y optimizadores, que obtuvieron como resultado el modelo que se muestra en la figura 2:

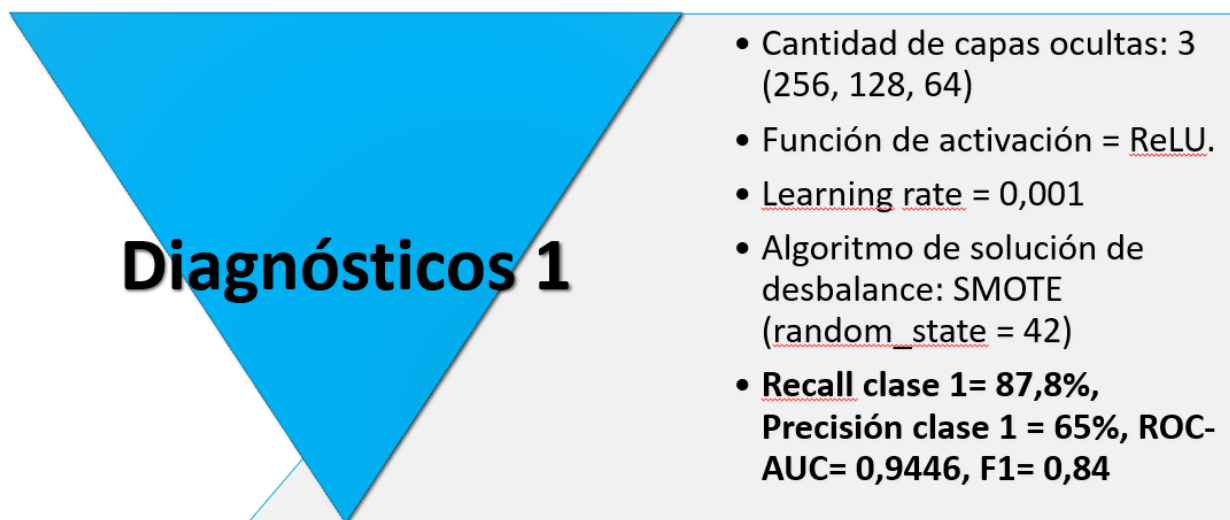


Figura 2. Configuración final del modelo

Evaluación y explicabilidad

El modelo se evaluó con validación cruzada estratificada (10 particiones) y pruebas de campo con datos no vistos que corresponden a Casos reales de pacientes ingresados en UCI con fecha de ingreso entre finales del año 2024 y el actual año 2025.

Las métricas principales fueron sensibilidad (recall), AUC-ROC, F1-score y precisión. Para mejorar la interpretabilidad, se implementó LIME (Ribeiro et al., 2016), generando explicaciones locales para cada predicción.

Construcción del Prototipo

El desarrollo se realizó en Python, utilizando bibliotecas como scikit-learn, imbalanced-learn, LIME y Streamlit (Khorasani et al., 2022) para la interfaz web. El pipeline integral incluyó preprocesamiento, balanceo, entrenamiento y explicabilidad. La figura 3 muestra una arquitectura funcional de la aplicación.

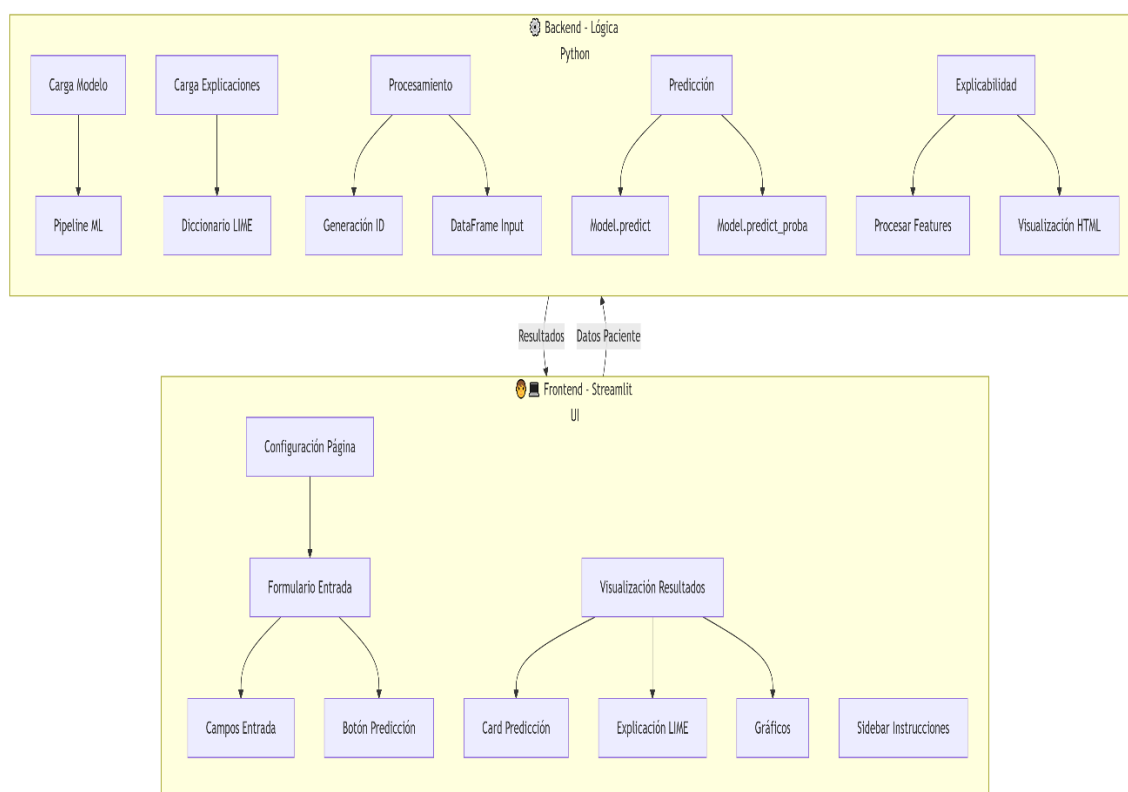


Figura 3. Arquitectura funcional de la aplicación WWW

RESULTADOS Y DISCUSIÓN

Desempeño del modelo MLP

El modelo MLP optimizado, con arquitectura de tres capas ocultas (256, 128 y 64 neuronas) y función de activación ReLU (Bengio et al., 2021), alcanzó los siguientes resultados durante el entrenamiento:

- Sensibilidad (clase mortalidad): 87,8%
- Precisión (clase mortalidad): 65%
- AUC-ROC: 0,9446
- F1-score: 0,75
- Precisión global: 88%

Pruebas de campo

En una base externa de 215 pacientes (41 fallecidos), el modelo mantuvo un rendimiento robusto:

- Sensibilidad: 85,37%
- AUC-ROC: 0,93
- F1-score: 0,70
- Precisión global: 86%

La matriz de confusión mostró 35 verdaderos positivos, 150 verdaderos negativos, 24 falsos positivos y solo 6 falsos negativos, destacando la capacidad del modelo para minimizar omisiones críticas.

Interpretabilidad con LIME

La integración de LIME permitió explicar las predicciones a nivel local, identificando las variables más influyentes en cada caso. Por ejemplo, en un paciente con alta probabilidad de supervivencia (figura 3), la estancia hospitalaria (stay) mostró la mayor contribución positiva, mientras que la edad y el APACHE actuaron como factores protectores. La métrica LSE (Local Sufficiency Explanation) (Ortigossa et al., 2024) superó 0,9 en la mayoría de los casos, indicando alta coherencia entre el modelo global y las explicaciones locales (Barredo Arrieta et al., 2019).

Para una correcta implementación del software de gestión de la producción se preparó una serie de instrucciones de trabajo donde se explica con imágenes paso a paso las acciones a realizar en cada puesto de trabajo. A continuación, se detallan los documentos elaborados, los cuales forman parte de la documentación del sistema de gestión de calidad implementado en la empresa basado en la norma ISO 13485 para la producción de dispositivos médicos.

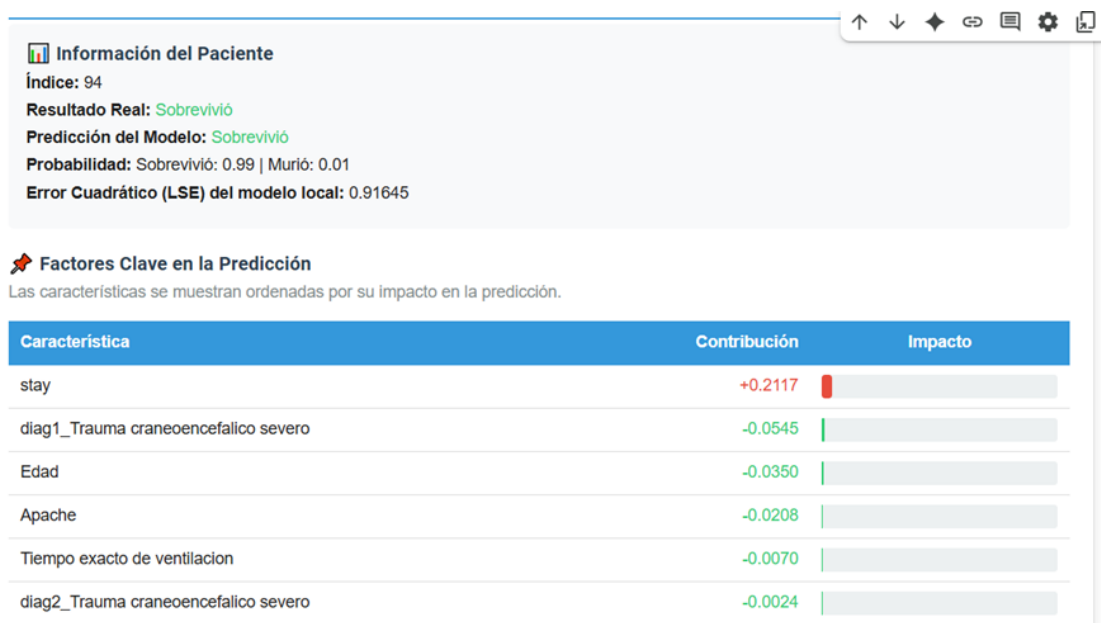


Figura 3. Resultados de aplicar LIME a la predicción de un paciente real

Comparación con modelos alternativos

El MLP superó consistentemente a KNN y Regresión Logística en sensibilidad y AUC-ROC, especialmente en conjuntos con diagnósticos codificados de forma detallada (Figura 5). Esto respalda su capacidad para capturar interacciones no lineales en datos clínicos complejos.

Categoría-diagnóstico-especificación 1					
Modelo	F1	ROC-AUC	Accuracy	Precisión	Recall
KNN	0,60	0,8637	0,79	0,49	0,79
Regresión Logística	0,51	0,8007	0,75	0,42	0,66
MLP	0,59	0,8743	0,78	0,46	0,81

Categoría-diagnóstico-especificación 2					
Modelo	F1	ROC-AUC	Accuracy	Precisión	Recall
KNN	0,58	0,8491	0,78	0,47	0,78
Regresión Logística	0,51	0,8007	0,75	0,42	0,66
MLP	0,57	0,8582	0,76	0,44	0,80

Diagnóstico 1					
Modelo	F1	ROC-AUC	Accuracy	Precisión	Recall
KNN	0,62	0,8678	0,81	0,51	0,78
Regresión Logística	0,53	0,8208	0,77	0,44	0,67
MLP	0,64	0,8820	0,83	0,54	0,78

Diagnóstico 2					
Modelo	F1	ROC-AUC	Accuracy	Precisión	Recall
KNN	0,57	0,8482	0,77	0,45	0,79
Regresión Logística	0,51	0,7782	0,72	0,39	0,73
MLP	0,57	0,8326	0,76	0,44	0,81

Figura 5. Comparación del desempeño de los algoritmos KNN, Regresión Logística y MLP en las diferentes bases de casos estudiadas

Limitaciones y consideraciones clínicas

El estudio presenta limitaciones derivadas principalmente de las características de la base de datos del Hospital Universitario "Arnaldo Milián Castro". Con 3.373 casos válidos tras preprocesamiento, el volumen de datos resulta insuficiente para entrenar arquitecturas de aprendizaje profundo más complejas (Johnson et al., 2016). Además, las inconsistencias en registros -valores faltantes en APACHE II, errores en tiempos de estancia y falta de estandarización diagnóstica- limitan la capacidad predictiva y generalización del modelo.

Clínicamente, se observa un trade-off relevante: el modelo alcanza alta sensibilidad (87,8%) pero precisión moderada (65%), generando falsos positivos que pueden llevar a sobreutilización de recursos en UCI. Esta compensación es aceptable en contextos donde priorizar la detección temprana de riesgo, pero requiere consideración en implementación clínica.

El uso complementario de bases de datos públicas como MIMIC-III representaría un aporte significativo (Johnson et al., 2016). Su volumen sustancial (>40,000 pacientes), diversidad poblacional y estandarización permitirían entrenar modelos más complejos y generalizables. MIMIC-III incluye variables clínicas adicionales (signos vitales secuenciales, puntuaciones SOFA diarias) no disponibles en la base local pero con valor pronóstico demostrado, que podrían mejorar el rendimiento predictivo.

La interpretabilidad, aunque mejorada con LIME(Ribeiro et al., 2016), sigue siendo limitada para un modelo MLP de "caja negra". Futuras iteraciones podrían incorporar SHAP o mecanismos de atención para explicaciones más completas (Došilović et al., 2018).

CONCLUSIONES

El sistema inteligente desarrollado, basado en una red neuronal MLP y técnicas de preprocesamiento robustas, demostró un alto desempeño en la predicción de mortalidad en UCI, con sensibilidad superior al 85% y AUC-ROC de 0,93. La integración de LIME (Ribeiro et al., 2016) mejora la transparencia del modelo, facilitando su adopción en entornos clínicos reales.

Los resultados validan el potencial de los enfoques de IA para apoyar la toma de decisiones en cuidados intensivos, especialmente en contextos con datos limitados y desbalanceados (Fernández et al., 2018). Futuras investigaciones podrían explorar el uso de técnicas de ensemble, la incorporación de variables temporales y la validación multicéntrica para fortalecer la generalización del modelo. Particularmente, la integración de la base de datos pública MIMIC-III (Medical Information Mart for Intensive Care III) representaría una oportunidad significativa para superar las limitaciones de volumen y calidad de los datos locales (Johnson et al., 2016). El uso de MIMIC-III, con sus más de 40.000 registros de pacientes de UCI y variables clínicas estandarizadas, permitiría entrenar arquitecturas más complejas, mejorar la capacidad de generalización del modelo a diferentes contextos hospitalarios.

AGRADECIMIENTOS

A los profesionales médicos del Hospital Universitario "Arnaldo Milián Castro" por su colaboración en la validación clínica de los datos. A los miembros del Proyecto Territorial: Plataforma de servicios de ayuda a la toma de decisiones en Unidades de Atención al Grave.

REFERENCIAS

- Azar, A. T., Elshazly, H. I., Hassanien, A. E., & Elkorany, A. M. (2014). A random forest classifier for lymph diseases. *Computer methods and programs in biomedicine*, 113(2), 465-473.
- Barredo Arrieta, A., Tabik, S., García López, S., Molina Cabrera, D., Herrera Triguero, F., & Díaz Rodríguez, N. A. (2019). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. <https://digibug.ugr.es/handle/10481/77982>
- Belciug, S., & Gorunescu, F. (2020). Era of Intelligent Systems in Healthcare. En S. Belciug & F. Gorunescu, *Intelligent Decision Support Systems—A Journey to Smarter Healthcare* (Vol. 157, pp. 1-55). Springer International Publishing. https://doi.org/10.1007/978-3-030-14354-1_1
- Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM*, 64(7), 58-65. <https://doi.org/10.1145/3448250>
- Došilović, F. K., Brčić, M., & Hlupić, N. (2018). Explainable artificial intelligence: A survey. 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO), 0210-0215. <https://ieeexplore.ieee.org/abstract/document/8400040/>
- Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from Imbalanced Data Sets*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-98074-4>

- Harnecker Fontecilla, M. B. (2017). Diseño y construcción de un modelo de predicción de mortalidad en pacientes adultos mayores hospitalizados en Unidad de Cuidados Intensivos del Hospital Clínico de la Universidad de Chile.
- Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), 160035. <https://doi.org/10.1038/sdata.2016.35>
- Khorasani, M., Abdou, M., & Hernández Fernández, J. (2022). Getting started with streamlit. En *Web Application Development with Streamlit: Develop and Deploy Secure and Scalable Web Applications to the Cloud Using a Pure Python Framework* (pp. 1-30). Springer.
- Nistal-Nuño, B. (2021). Artificial intelligence forecasting mortality at an intensive care unit and comparison to a logistic regression system. *Einstein (São Paulo)*, 19, eAO6283. https://doi.org/10.31744/einstein_journal/2021AO6283
- Ortigosa, E. S., Gonçalves, T., & Nonato, L. G. (2024). EXplainable artificial intelligence (XAI)—From theory to methods and applications. *IEEE Access*.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). «Why Should I Trust You?»: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. <https://doi.org/10.1145/2939672.2939778>
- Sánchez Martín, D. (2024). Algoritmos de selección de características en aprendizaje automático. <https://gredos.usal.es/handle/10366/164911>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 12(3), 109-124.
- Vieira, P. M., & Rodrigues, F. (2024). An automated approach for binary classification on imbalanced data. *Knowledge and Information Systems*, 66(5), 2747-2767. <https://doi.org/10.1007/s10115-023-02046-7>
- Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295-316. <https://doi.org/10.1016/j.neucom.2020.07.061>

Copyright © 2025, Autores: Vicente González, María del Rocío., García Lorenzo, María Matilde, Fernández Fleites, Rafael, Vicente Rodríguez, Roberto.



Esta obra está bajo una licencia de Creative Commons Atribución-No Comercial 4.0 Internacional