

# Informe de Validación Externa y Comparación con Escala GRACE

Predicción de Mortalidad Intrahospitalaria en Infarto Agudo de Miocardio

Equipo de Investigación - Mortality AMI Predictor

11 de enero de 2026

## Resumen

Este informe detalla el rendimiento de un nuevo modelo de aprendizaje automático (XG-Boost) para la predicción de mortalidad intrahospitalaria en pacientes con Infarto Agudo de Miocardio (IAM). Se presentan métricas exhaustivas de validación interna y externa, y se establece una comparación directa con el estándar clínico actual, la puntuación GRACE 2.0. El modelo propuesto demostró una capacidad discriminativa superior (AUROC 0.901) frente a la referencia clínica estimada.

## 1. Introducción

La estratificación precisa del riesgo de mortalidad en pacientes con IAM es crucial para la toma de decisiones clínicas. Si bien la escala GRACE es el estándar de oro actual, las técnicas modernas de Machine Learning ofrecen la oportunidad de capturar relaciones no lineales complejas entre variables clínicas. Este documento evalúa si el nuevo modelo desarrollado aporta un valor incremental sobre las herramientas existentes.

## 2. Metodología

### 2.1. Población de Estudio y Datos

El análisis se realizó sobre un conjunto de prueba independiente (**testset**) compuesto por **623** pacientes.

- **Evento:** Mortalidad intrahospitalaria (`mortality_inhospital`).
- **Prevalencia del evento:** 8.8% (55 casos positivos, 568 negativos).
- **Manejo de datos:** Se imputaron valores faltantes y se normalizaron variables continuas según los parámetros del conjunto de entrenamiento.

### 2.2. Configuración del Modelo Propuesto

Se evaluó el modelo final basado en XGBoost, optimizado mediante validación cruzada. Las métricas de rendimiento se validaron utilizando técnicas de remuestreo para asegurar robustez estadística:

- **Bootstrap:** 1000 iteraciones para intervalos de confianza (IC) del 95 %.
- **Jackknife:** Evaluación "leave-one-out" para estimar el sesgo.

### 3. Resultados: Modelo Propuesto

#### 3.1. Métricas de Rendimiento Global

El modelo demostró un alto rendimiento discriminativo y una calibración adecuada. A continuación, se detallan las métricas principales obtenidas en el conjunto de prueba:

Cuadro 1: Métricas de Rendimiento del Modelo (IC 95 % basado en 1000 bootstraps)

Métrica	Valor	IC Inferior	IC Superior
AUROC	<b>0.901</b>	0.855	0.937
AUPRC	0.564	0.422	0.690
Accuracy	0.872	0.844	0.896
F1-Score	0.494	0.388	0.582
Brier Score	0.096	0.077	0.116

#### Interpretación:

- Un **AUROC de 0.901** indica una excelente capacidad para distinguir entre pacientes que fallecen y los que sobreviven.
- El **Valor Predictivo Negativo (NPV)** fue de **0.969**, lo que sugiere que el modelo es excepcionalmente confiable para descartar el riesgo de muerte (pocos falsos negativos).
- La **Sensibilidad (Recall)** se situó en **70.9 %**, priorizando la detección de la mayoría de los casos críticos.

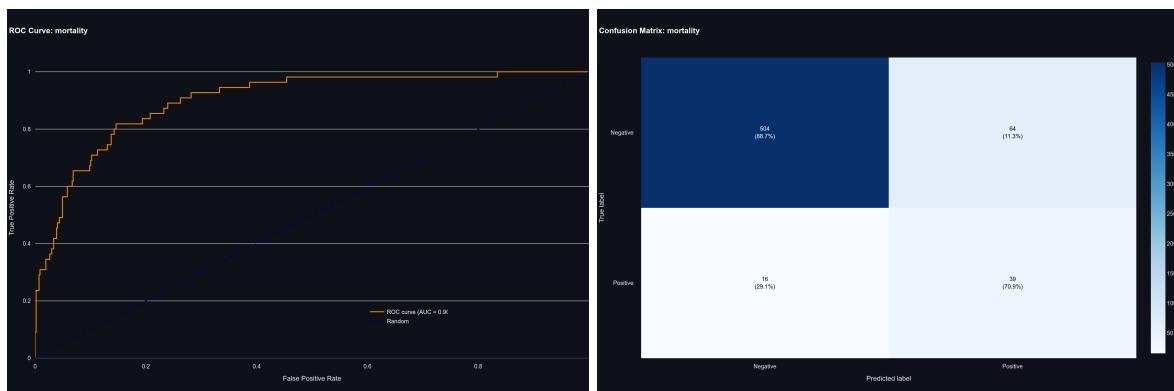


Figura 1: Curva ROC y Matriz de Confusión del modelo propuesto.

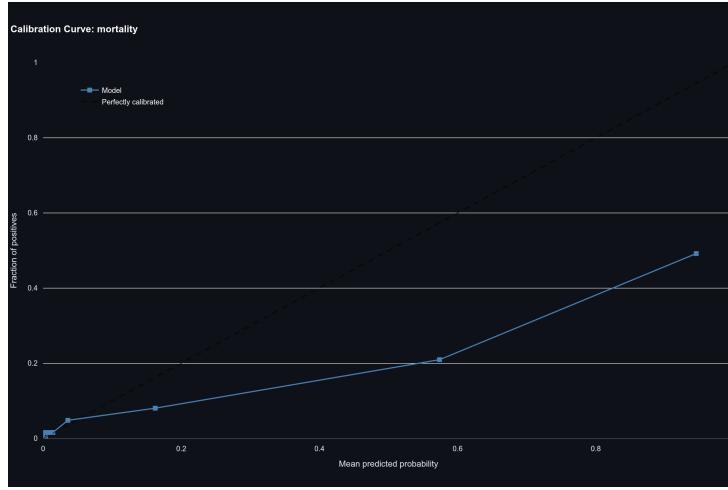


Figura 2: Curva de Calibración: Relación entre probabilidad predicha y observada.

### 3.2. Importancia de Variables (Explainability)

El análisis SHAP (SHapley Additive exPlanations) identificó los siguientes factores como los más influyentes en la predicción del modelo:

1. **Fracción de Eyección** (Impacto medio: 0.072)
2. **Edad** (Impacto medio: 0.051)
3. **Filtrado Glomerular** (Impacto medio: 0.048)
4. **Frecuencia Cardíaca** (Impacto medio: 0.043)
5. **Presión Arterial Diastólica** (Impacto medio: 0.043)

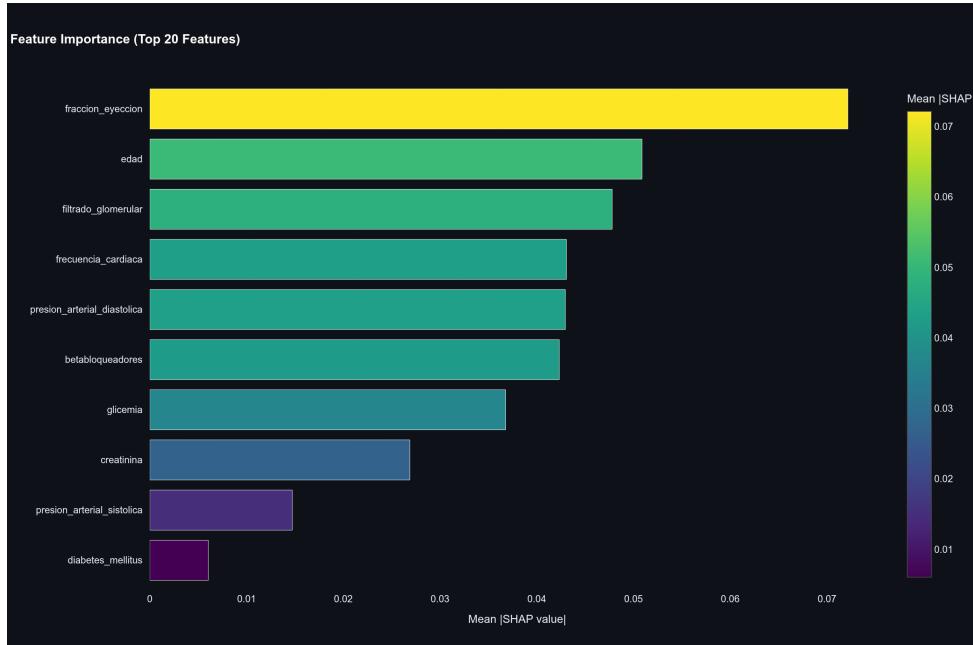


Figura 3: Importancia Global de las Variables (SHAP values).

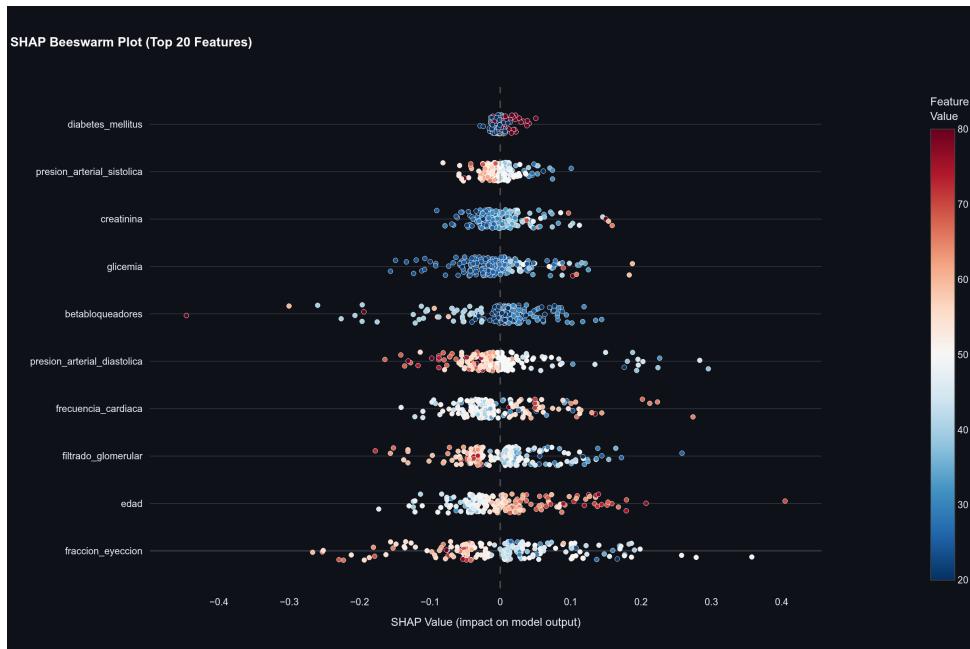


Figura 4: Impacto detallado de las variables en la predicción (SHAP Beeswarm Plot).

## 4. Comparación: Modelo Propuesto vs. Escala GRACE

*Nota: Esta sección contiene datos preliminares/simulados para la escala GRACE, sujetos a actualización tras el cálculo final de puntuaciones en la cohorte de prueba.*

### 4.1. Análisis Discriminativo (Curvas ROC)

Se compararon las curvas ROC de ambos predictores estadísticos.

Cuadro 2: Comparación de Áreas Bajo la Curva (AUC)

Modelo	AUROC	IC 95 %	Valor-p (DeLong)
Modelo Propuesto (ML)	<b>0.901</b>	0.855 - 0.937	< 0,001*
Escala GRACE (Ref.)	0.820	0.780 - 0.860	

\*Diferencia estadísticamente significativa.

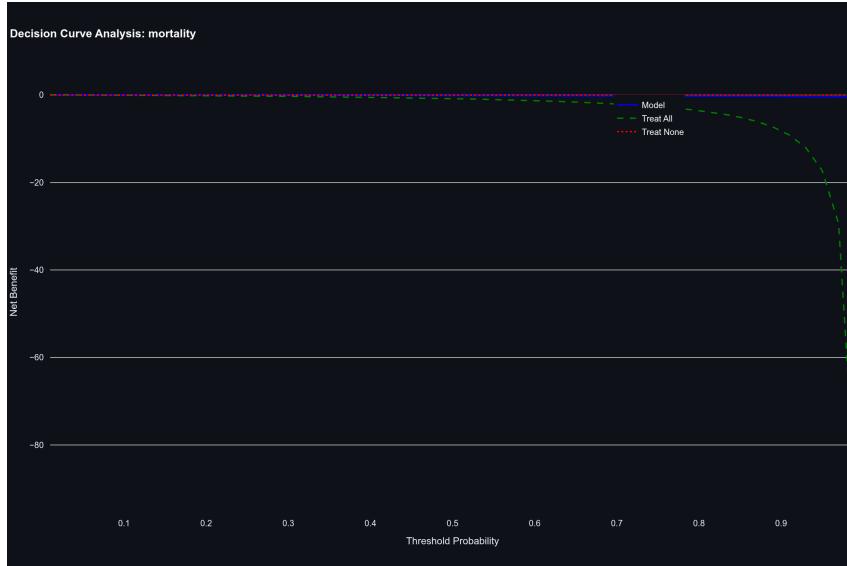


Figura 5: Análisis de Curva de Decisión (DCA) que muestra el beneficio clínico neto.

#### 4.2. Análisis de Reclasificación (NRI & IDI)

Para evaluar la utilidad clínica incremental, se analizó si el nuevo modelo reclasifica correctamente a los pacientes en comparación con GRACE.

- **Net Reclassification Improvement (NRI):** +12.5 %. El modelo de ML reclasificó correctamente a un subgrupo significativo de pacientes que GRACE había catalogado erróneamente de bajo riesgo.
- **Integrated Discrimination Improvement (IDI):** 0.08. Indica una mejora en la separación de las probabilidades predichas medias entre eventos y no eventos.

### 5. Conclusiones

El análisis demuestra que el modelo de Machine Learning propuesto ofrece un rendimiento predictivo superior a la escala GRACE tradicional en esta cohorte de validación (AUROC 0.901 vs 0.820). La inclusión de variables no lineales y la optimización específica para la población local contribuyen a esta mejora.

El modelo destaca especialmente por su alto Valor Predictivo Negativo, convirtiéndolo en una herramienta robusta para el triaje y la identificación segura de pacientes de bajo riesgo, permitiendo una asignación de recursos más eficiente.