

# Predicción de Mortalidad Intrahospitalaria por Infarto Agudo de Miocardio mediante Aprendizaje Automático

Informe de Investigación

Basado en el Registro Cubano de Infarto Agudo de Miocardio

(RECUIMA)

## **Equipo 7**

Richard Alejandro Matos Arderí

Abel Ponce González

Abraham Romero Imbert

Eveliz Espinaco Milián

Michell Viu Ramirez

Ernesto Abreu Peraza

Eduardo Brito Labrada

*Curso de Machine Learning*

*Universidad de Ciencias Médicas de La Habana*

12 de enero de 2026

## Resumen

**Contexto:** El infarto agudo de miocardio (IAM) representa una de las principales causas de mortalidad cardiovascular a nivel mundial. La predicción temprana de la mortalidad intrahospitalaria permite optimizar la estratificación de riesgo y la asignación de recursos terapéuticos.

**Objetivo:** Desarrollar y validar modelos de aprendizaje automático para predecir la mortalidad intrahospitalaria en pacientes con IAM utilizando datos del Registro Cubano de Infarto Agudo de Miocardio (RECUIMA).

**Métodos:** Se analizaron **3.112** pacientes con diagnóstico de IAM registrados en el RECUIMA. Se implementaron múltiples algoritmos de clasificación incluyendo Regresión Logística, Random Forest, XGBoost, LightGBM y K-Nearest Neighbors. El rendimiento se evaluó mediante validación cruzada estratificada y métricas de discriminación (AUROC, AUPRC), calibración y utilidad clínica. Se desarrollaron dos enfoques: un modelo reducido (10 variables) comparable con la escala GRACE y un modelo extendido (57 variables) como propuesta original.

**Resultados:** El modelo con mejor rendimiento fue **XGBoost con el conjunto extendido de variables**, alcanzando un AUROC de **0,938 (IC 95 %: 0,884–0,977)** en el conjunto de prueba. El modelo reducido, comparable con GRACE, obtuvo un AUROC de **0,901 (IC 95 %: 0,855–0,937)**. Las variables más predictivas identificadas mediante análisis SHAP fueron **fracción de eyección, edad, filtrado glomerular, frecuencia cardíaca y presión arterial diastólica**. Ambos modelos superaron el rendimiento de las escalas clásicas GRACE (AUROC estimado: 0,820) en esta población.

**Conclusiones:** Los modelos de aprendizaje automático demuestran capacidad superior para predecir mortalidad intrahospitalaria por IAM en comparación con escalas tradicionales. La interpretabilidad mediante SHAP facilita la adopción clínica al identificar los factores de riesgo más relevantes.

**Palabras clave:** infarto agudo de miocardio, mortalidad intrahospitalaria, aprendizaje automático, predicción de riesgo, SHAP, XGBoost

## Abstract

**Background:** Acute myocardial infarction (AMI) remains one of the leading causes of cardiovascular mortality worldwide. Early prediction of in-hospital mortality enables optimal risk stratification and therapeutic resource allocation.

**Objective:** To develop and validate machine learning models for predicting in-hospital mortality in AMI patients using data from the Cuban Registry of Acute Myocardial Infarction (RECUIMA).

**Methods:** We analyzed **3,112** patients diagnosed with AMI from RECUIMA. Multiple classification algorithms were implemented, including Logistic Regression, Random Forest, XGBoost, LightGBM and K-Nearest Neighbors. Performance was evaluated using stratified cross-validation with discrimination metrics (AUROC, AUPRC), calibration, and clinical utility. Two approaches were developed: a reduced model (10 variables) comparable with GRACE score and an extended model (57 variables) as our original proposal.

**Results:** The best-performing model was **XGBoost with extended variable set**, achieving an AUROC of **0.938 (95 %CI: 0.884–0.977)** on the test set. The reduced model, comparable with GRACE, achieved an AUROC of **0.901 (95 %CI: 0.855–0.937)**. The most predictive variables identified through SHAP analysis were **ejection fraction, age, glomerular filtration rate, heart rate, and diastolic blood pressure**. Both models outperformed traditional GRACE (estimated AUROC: 0.820) scores in this population.

**Conclusions:** Machine learning models demonstrate superior capability for predicting in-hospital mortality from AMI compared to traditional scores. Interpretability through SHAP facilitates clinical adoption by identifying the most relevant risk factors.

**Keywords:** acute myocardial infarction, in-hospital mortality, machine learning, risk prediction, SHAP, XGBoost

# Índice

<b>1. Introducción</b>	<b>15</b>
1.1. Contexto y Motivación . . . . .	15
1.2. Escalas de Riesgo Tradicionales . . . . .	15
1.3. Limitaciones de los Enfoques Tradicionales . . . . .	16
1.4. Aprendizaje Automático en Predicción Clínica . . . . .	16
1.5. Objetivos del Estudio . . . . .	16
1.5.1. Objetivo General . . . . .	16
1.5.2. Objetivos Específicos . . . . .	17
1.6. Estructura del Informe . . . . .	17
<b>2. Estado del Arte</b>	<b>18</b>
2.1. Modelos Clásicos de Predicción de Mortalidad en IAM . . . . .	18
2.1.1. Escala GRACE . . . . .	18
2.1.2. Escala TIMI . . . . .	18
2.1.3. Otras Escalas Contemporáneas . . . . .	19
2.2. Limitaciones de los Modelos Tradicionales . . . . .	19
2.3. Aprendizaje Automático en Predicción Cardiovascular . . . . .	20
2.3.1. Evolución Histórica . . . . .	20
2.3.2. Estudios Relevantes . . . . .	20
2.3.3. Comparación de Algoritmos . . . . .	21
2.4. Predicción de Arritmias Ventriculares . . . . .	21
2.5. Desafíos y Vacíos en la Literatura . . . . .	21
2.6. Oportunidades de Investigación . . . . .	22
2.7. Marco Teórico del Presente Estudio . . . . .	22
<b>3. Descripción del Conjunto de Datos</b>	<b>22</b>
3.1. Fuente de Datos . . . . .	22
3.2. Tamaño Muestral y Estructura . . . . .	23
3.3. Categorización de Variables . . . . .	23
3.3.1. Variables Demográficas y Administrativas . . . . .	23
3.3.2. Antecedentes Patológicos . . . . .	23
3.3.3. Presentación Clínica . . . . .	24
3.3.4. Variables Electrocardiográficas . . . . .	24
3.3.5. Biomarcadores de Laboratorio . . . . .	24
3.3.6. Tratamientos y Procedimientos . . . . .	25
3.3.7. Variables de Imagen . . . . .	25
3.3.8. Complicaciones Intrahospitalarias . . . . .	26

3.3.9. Variables Outcome . . . . .	26
3.4. Consideraciones sobre Calidad de Datos . . . . .	26
3.4.1. Datos Faltantes . . . . .	26
3.4.2. Valores Atípicos Identificados . . . . .	28
3.5. Resumen Descriptivo Preliminar . . . . .	29
<b>4. Metodología</b>	<b>29</b>
4.1. Diseño del Estudio . . . . .	29
4.2. Preprocesamiento de Datos . . . . .	29
4.2.1. Tratamiento de Datos Faltantes . . . . .	29
4.2.2. Tratamiento de Valores Atípicos . . . . .	30
4.2.3. Codificación de Variables Categóricas . . . . .	30
4.2.4. Transformación y Normalización . . . . .	31
4.3. Ingeniería de Características . . . . .	31
4.3.1. Creación de Variables Derivadas . . . . .	31
4.3.2. Selección de Características . . . . .	31
4.4. División de Datos . . . . .	32
4.5. Manejo del Desbalance de Clases . . . . .	32
4.6. Algoritmos de Aprendizaje Automático . . . . .	32
4.6.1. Modelos Evaluados . . . . .	32
4.6.2. Optimización de Hiperparámetros . . . . .	33
4.7. Métricas de Evaluación . . . . .	33
4.7.1. Discriminación . . . . .	33
4.7.2. Calibración . . . . .	34
4.7.3. Utilidad Clínica . . . . .	34
4.8. Análisis de Explicabilidad . . . . .	34
4.8.1. Explicabilidad Global . . . . .	34
4.8.2. Explicabilidad Local . . . . .	34
4.9. Validación del Modelo . . . . .	35
4.10. Herramientas Tecnológicas . . . . .	35
4.11. Reproducibilidad . . . . .	35
<b>5. Análisis Exploratorio de Datos</b>	<b>36</b>
5.1. Distribución de la Variable Objetivo . . . . .	36
5.2. Análisis de Variables Demográficas . . . . .	36
5.2.1. Distribución de Edad . . . . .	37
5.2.2. Distribución por Sexo . . . . .	38
5.3. Análisis de Antecedentes Patológicos . . . . .	38
5.4. Análisis de Variables Clínicas de Presentación . . . . .	39
5.4.1. Signos Vitales al Ingreso . . . . .	39

5.4.2. Clasificación Killip al Ingreso . . . . .	40
5.5. Análisis de Biomarcadores . . . . .	40
5.5.1. Biomarcadores Renales y Metabólicos . . . . .	41
5.6. Análisis de Correlaciones . . . . .	42
5.7. Análisis de Tratamientos y Procedimientos . . . . .	42
5.8. Resumen de Hallazgos del EDA . . . . .	43
<b>6. Modelado Predictivo</b>	<b>44</b>
6.1. Estrategia General de Modelado . . . . .	44
6.2. Modelos Baseline . . . . .	44
6.2.1. Regresión Logística . . . . .	44
6.3. Algoritmos de Aprendizaje Automático Evaluados . . . . .	45
6.3.1. Random Forest . . . . .	45
6.3.2. XGBoost . . . . .	45
6.3.3. LightGBM . . . . .	46
6.3.4. Redes Neuronales . . . . .	46
6.3.5. AutoML y Neural Architecture Search . . . . .	46
6.4. Proceso de Optimización . . . . .	47
6.4.1. Método de Búsqueda . . . . .	47
6.5. Comparación de Modelos en Validación . . . . .	48
6.6. Selección del Modelo Final . . . . .	49
6.6.1. Criterios de Selección . . . . .	49
6.6.2. Modelo Seleccionado . . . . .	50
6.7. Calibración del Modelo . . . . .	50
6.7.1. Curva de Calibración . . . . .	51
6.7.2. Curva de Aprendizaje . . . . .	51
<b>7. Resultados</b>	<b>51</b>
7.1. Métricas de Discriminación . . . . .	52
7.1.1. Comparación de Ambos Enfoques . . . . .	52
7.1.2. Modelo Reducido: Detalles (Comparable con GRACE) . . . . .	53
7.1.3. Modelo Extendido: Detalles (Propuesta Principal) . . . . .	53
7.1.4. Curva ROC . . . . .	54
7.1.5. Matriz de Confusión . . . . .	55
7.2. Métricas de Calibración . . . . .	55
7.2.1. Curva de Calibración . . . . .	55
7.2.2. Métricas de Calibración . . . . .	56
7.3. Comparación con Modelos de Referencia . . . . .	56
7.3.1. Tests de Significancia Estadística . . . . .	56
7.4. Análisis de Utilidad Clínica . . . . .	56

7.4.1.	Decision Curve Analysis . . . . .	57
7.4.2.	Net Reclassification Index (NRI) . . . . .	57
7.5.	Resumen de Resultados . . . . .	58
7.6.	Comparaciones Visuales entre Modelos . . . . .	58
7.7.	Validación con Dataset sin Fuga de Datos . . . . .	60
7.7.1.	Métricas del Modelo sin Variables con Potencial Fuga . . . . .	61
7.7.2.	Curva ROC del Modelo Validado . . . . .	62
7.7.3.	Matriz de Confusión del Modelo Validado . . . . .	62
7.7.4.	Importancia de Variables (Sin Fuga de Datos) . . . . .	63
7.7.5.	Curvas de Aprendizaje . . . . .	63
7.7.6.	Comparación Estadística entre Modelos . . . . .	64
7.7.7.	Análisis de Utilidad Clínica . . . . .	65
7.7.8.	Validación Bootstrap y Jackknife . . . . .	66
<b>8.</b>	<b>Explicabilidad del Modelo</b>	<b>66</b>
8.1.	Importancia Global de Variables . . . . .	66
8.1.1.	Feature Importance Nativa . . . . .	67
8.1.2.	SHAP Summary Plot . . . . .	67
8.1.3.	SHAP Bar Plot . . . . .	69
8.2.	Efectos Parciales de Variables Individuales . . . . .	69
8.2.1.	SHAP Dependence Plots . . . . .	69
8.2.2.	Interpretación de Efectos . . . . .	69
8.3.	Interacciones entre Variables . . . . .	70
8.3.1.	SHAP Interaction Values . . . . .	70
8.3.2.	Interacciones Clínicamente Relevantes . . . . .	70
8.4.	Explicaciones Locales (Individuales) . . . . .	71
8.4.1.	SHAP Force Plots . . . . .	71
8.4.2.	SHAP Waterfall Plots . . . . .	72
8.5.	Validación Clínica de la Explicabilidad . . . . .	73
8.6.	Consistencia con Conocimiento Clínico . . . . .	74
8.7.	Implicaciones para la Práctica Clínica . . . . .	74
8.8.	Análisis SHAP del Modelo sin Fuga de Datos . . . . .	75
8.8.1.	Importancia Global de Variables (Sin Fuga) . . . . .	75
8.8.2.	Hallazgos Clave del Análisis SHAP sin Fuga . . . . .	75
8.9.	Integración en la Herramienta Dashboard . . . . .	76
<b>9.</b>	<b>Discusión</b>	<b>77</b>
9.1.	Resumen de Hallazgos Principales . . . . .	77
9.2.	Comparación con la Literatura . . . . .	77
9.2.1.	Rendimiento Comparativo . . . . .	78

9.2.2. Variables Predictoras . . . . .	78
9.3. Fortalezas del Estudio . . . . .	79
9.4. Limitaciones del Estudio . . . . .	79
9.4.1. Limitaciones Metodológicas . . . . .	79
9.4.2. Limitaciones de los Datos . . . . .	80
9.4.3. Limitaciones del Modelo . . . . .	80
9.5. Implicaciones Clínicas . . . . .	80
9.5.1. Potencial de Implementación . . . . .	81
9.5.2. Beneficios Potenciales . . . . .	81
9.5.3. Consideraciones de Implementación . . . . .	81
9.6. Consideraciones Éticas . . . . .	81
<b>10. Conclusiones</b>	<b>82</b>
10.1. Conclusiones Principales . . . . .	82
10.2. Respuesta a los Objetivos . . . . .	83
10.2.1. Objetivo General . . . . .	83
10.2.2. Objetivos Específicos . . . . .	83
10.3. Contribuciones del Estudio . . . . .	84
10.4. Mensaje Clave . . . . .	85
10.5. Recomendaciones . . . . .	85
10.5.1. Para Investigadores . . . . .	85
10.5.2. Para Clínicos . . . . .	85
10.5.3. Para Gestores de Salud . . . . .	85
10.6. Reflexión Final . . . . .	86
<b>Referencias Bibliográficas</b>	<b>87</b>



## Índice de figuras

1.	Distribución de la variable objetivo (mortalidad intrahospitalaria). Se observa el marcado desbalance entre las clases: la mayoría de pacientes sobrevivieron (clase 0), mientras que una minoría falleció (clase 1), característico de estudios de mortalidad cardiovascular. . . . .	36
2.	Distribución de edad en la población de estudio. El histograma muestra una distribución aproximadamente normal con ligera asimetría hacia edades mayores, con mayor concentración de pacientes entre 55 y 75 años. . . . .	37
3.	Análisis comparativo de edad. Los diagramas de caja y violín revelan la presencia de valores atípicos en edades extremas y confirman la distribución unimodal de la variable. . . . .	37
4.	Distribución por sexo en la cohorte de estudio. Se observa predominio del sexo masculino, consistente con la epidemiología del IAM donde los hombres presentan mayor incidencia, especialmente en edades más tempranas. . . . .	38
5.	Distribución de los principales factores de riesgo cardiovascular. La hipertensión arterial presenta alta prevalencia en la cohorte, mientras que la diabetes mellitus afecta aproximadamente a un tercio de los pacientes. . . . .	38
6.	Distribución de antecedentes patológicos adicionales. El tabaquismo constituye un factor de riesgo modificable frecuente, mientras que la insuficiencia renal crónica, aunque menos prevalente, se asocia con peor pronóstico. . . . .	38
7.	Análisis de antecedentes patológicos según desenlace. Los violin plots permiten visualizar la distribución de estos factores de riesgo estratificados por mortalidad intrahospitalaria. . . . .	39
8.	Distribución de presión arterial al ingreso según desenlace. Se observa que los pacientes fallecidos tienden a presentar valores más bajos de presión arterial, lo cual puede indicar compromiso hemodinámico o shock cardiogénico. . . . .	39
9.	Distribución de frecuencia cardíaca según desenlace. La frecuencia cardíaca elevada al ingreso se asocia con mayor mortalidad, reflejando activación simpática compensatoria en contexto de disfunción ventricular. . . . .	40
10.	Distribución de la clasificación Killip-Kimball. La mayoría de pacientes ingresaron en clase Killip I (sin insuficiencia cardíaca), mientras que un porcentaje menor presentó grados avanzados de disfunción ventricular (clases III-IV). . . . .	40

11.	Distribución de creatinina sérica. Se observa una distribución asimétrica positiva con valores atípicos en el rango superior, indicando presencia de pacientes con disfunción renal significativa. . . . .	41
12.	Análisis de creatinina según desenlace. Los niveles elevados de creatinina se asocian con mayor mortalidad, reflejando el impacto pronóstico de la disfunción renal en el contexto del IAM. . . . .	41
13.	Distribución de glucemia al ingreso. La hiperglucemia de estrés es frecuente en el IAM y se asocia con peor pronóstico, independientemente del estado diabético previo. . . . .	41
14.	Matriz de correlaciones entre variables numéricas del dataset. El mapa de calor permite identificar grupos de variables correlacionadas: se observan correlaciones esperadas entre variables hemodinámicas (presiones arteriales), entre marcadores de función renal, y entre variables relacionadas con la gravedad clínica. . . . .	42
15.	Análisis de tratamientos farmacológicos. El uso de aminas vasoactivas se asocia con mayor mortalidad (indicador de shock cardiogénico), mientras que los betabloqueadores muestran efecto cardioprotector. . . . .	43
16.	Análisis de complicaciones arrítmicas. Tanto la fibrilación ventricular como la taquicardia ventricular se asocian fuertemente con mortalidad intrahospitalaria, representando complicaciones eléctricas graves del IAM. . . . .	43
17.	Matrices de comparación estadística entre modelos. Izquierda: p-valores (pruebas pareadas). Derecha: Tamaño del efecto (d de Cohen). Valores de p cercanos a cero indican diferencias significativas. Tamaños de efecto $> 0,8$ se consideran grandes. . . . .	49
18.	Comparación directa entre Regresión Logística (baseline) y XGBoost. Se observa la mejora significativa en el rendimiento predictivo del modelo XGBoost. . . . .	49
19.	Curva de calibración del modelo reducido. La diagonal representa calibración perfecta. El modelo muestra buena calibración general con Brier Score de 0,096. . . . .	51
20.	Curva de aprendizaje del modelo XGBoost. Muestra la evolución del rendimiento (score) en función del tamaño del conjunto de entrenamiento. La convergencia de las curvas de entrenamiento y validación indica un buen balance sin overfitting significativo. . . . .	51
21.	Curva ROC del modelo reducido (10 variables) en conjunto de test. AU-ROC = 0,901 (IC 95 %: 0,855–0,937). La curva muestra el rendimiento del modelo XGBoost optimizado para comparación con escalas internacionales. . . . .	54

22.	Matriz de confusión del modelo reducido (umbral óptimo de Youden). VP: Verdaderos Positivos, VN: Verdaderos Negativos, FP: Falsos Positivos, FN: Falsos Negativos. . . . .	55
23.	Curva de calibración del modelo reducido. La línea diagonal representa calibración perfecta. El modelo muestra buena calibración con Brier Score = 0,096. . . . .	55
24.	Análisis de curva de decisión (DCA). El modelo muestra beneficio neto positivo sobre las estrategias de “tratar a todos” o “tratar a ninguno” en el rango de umbrales clínicamente relevantes (5 %–30 %). . . . .	57
25.	Comparación entre Random Forest y XGBoost. Aunque ambos modelos presentan rendimiento similar, XGBoost muestra mejor calibración y menor variabilidad. . . . .	59
26.	Comparación entre XGBoost y LightGBM. Los modelos de gradient boosting muestran rendimiento comparable, sin diferencias estadísticamente significativas. . . . .	59
27.	Comparación entre XGBoost estándar y XGBoost con balance de clases. El manejo del desbalance mediante scale_pos_weight no impacta significativamente el rendimiento final. . . . .	60
28.	Comparación entre K-Nearest Neighbors y XGBoost. XGBoost supera claramente al modelo KNN, demostrando la superioridad de los métodos de ensemble para este problema. . . . .	60
29.	Curva ROC del modelo XGBoost entrenado sin variables con potencial fuga de datos. AUROC = 0,896. El modelo mantiene excelente capacidad discriminativa utilizando únicamente predictores clínicamente válidos. . . . .	62
30.	Matriz de confusión del modelo sin fuga de datos. VN: 559 (98,4 %), FP: 9 (1,6 %), FN: 34 (61,8 %), VP: 21 (38,2 %). El modelo mantiene alta especificidad con precisión aceptable. . . . .	62
31.	Top 20 variables más importantes según valores SHAP medios absolutos en el modelo sin fuga de datos. Las variables más influyentes son: edad (0,019), fracción de eyección (0,017), glicemia (0,017), índice Killip (0,016) y presión arterial diastólica (0,015). Ninguna de estas variables propicia fuga de datos. . . . .	63
32.	Curva de aprendizaje de XGBoost (dataset sin fuga). Score final en validación: 0,908. Gap train-val: 0,092. La convergencia de las curvas indica generalización adecuada. . . . .	64
33.	Curva de aprendizaje de Random Forest (dataset sin fuga). Score final en validación: 0,896. Gap train-val: 0,104. . . . .	64

34.	Matriz de comparaciones estadísticas entre modelos. Izquierda: p-valores (verde = diferencia significativa). Derecha: tamaño del efecto (Cohen's d). XGBoost y XGBoost Balanced superan significativamente a LightGBM.	65
35.	Decision Curve Analysis del modelo sin fuga de datos. El modelo (línea azul) muestra beneficio neto positivo sobre la estrategia de "tratar a ninguno" en todo el rango de umbrales clínicamente relevantes, confirmando su utilidad clínica incluso sin variables potencialmente sesgadas. . . . .	65
36.	Distribución del AUROC mediante Bootstrap (n=1000, izquierda) y Jackknife (n=623, derecha). Los intervalos de confianza al 95 % confirman la estabilidad del rendimiento del modelo. . . . .	66
37.	Distribución del AUPRC mediante técnicas de remuestreo. La métrica muestra mayor variabilidad debido al desbalance de clases, pero mantiene valores aceptables. . . . .	66
38.	Importancia de características según el modelo XGBoost (gain). Las variables más importantes son: filtrado glomerular, fracción de eyección, edad, glicemia, presión arterial diastólica y creatinina. . . . .	67
39.	SHAP Summary Plot (Beeswarm) del modelo reducido. Cada punto representa un paciente. El color indica el valor de la variable (rojo = alto, azul = bajo). El eje X muestra el impacto en la predicción (valores SHAP positivos aumentan el riesgo de mortalidad). . . . .	67
40.	Importancia media de variables según valores SHAP para el modelo extendido (top 20 características). El valor SHAP medio absoluto indica la magnitud del impacto de cada variable en las predicciones. . . . .	69
41.	SHAP Summary Plot (Beeswarm) del modelo extendido (top 20 características). Mayor densidad de información sobre los predictores más relevantes del conjunto expandido de 57 variables. . . . .	69
42.	SHAP Force Plot – Ejemplo de predicción individual. Los factores en rojo aumentan el riesgo de mortalidad, los factores en azul lo disminuyen. La longitud de cada segmento indica la magnitud del efecto. . . . .	71
43.	SHAP Force Plot – Paciente del modelo extendido (muestra 156). Visualización de cómo cada variable contribuye a la predicción final de riesgo para este paciente específico. . . . .	72
44.	SHAP Waterfall Plot del modelo reducido – Caso representativo. Muestra la contribución de cada variable a la predicción final, partiendo del valor base (probabilidad media) hasta la predicción individual. . . . .	72
45.	SHAP Waterfall Plot del modelo extendido (muestra 0, base: 0,0416, predicción: 0,0009). Paciente de bajo riesgo donde múltiples factores protectores reducen la probabilidad predicha. . . . .	73

46. Importancia de variables según SHAP en el modelo sin fuga de datos. Las 5 variables más influyentes son: edad, fracción de eyección, glicemia, índice Killip y presión arterial diastólica — todas disponibles al ingreso del paciente y sin riesgo de sesgo. . . . . 75

## Índice de tablas

1. Variables incluidas en la escala GRACE . . . . .	18
2. Variables incluidas en la escala TIMI para IAM con elevación del ST . . . . .	19
3. Comparación de algoritmos de ML en predicción de mortalidad por IAM (revisión de literatura) . . . . .	21
4. Información general del conjunto de datos . . . . .	22
5. Resumen del tamaño muestral . . . . .	23
6. Variables demográficas y administrativas . . . . .	23
7. Variables de antecedentes patológicos personales . . . . .	23
8. Variables de presentación clínica al ingreso . . . . .	24
9. Variables electrocardiográficas . . . . .	24
10. Biomarcadores y pruebas de laboratorio . . . . .	24
11. Variables de tratamiento y procedimientos . . . . .	25
12. Variables de estudios de imagen . . . . .	25
13. Variables de complicaciones durante la hospitalización . . . . .	26
14. Variables de desenlace (outcome) . . . . .	26
15. Variables con >90 % de datos faltantes (excluidas) . . . . .	27
16. Variables con 20–90 % de datos faltantes . . . . .	27
17. Variables clínicas relevantes con <20 % de datos faltantes . . . . .	28
18. Características del diseño del estudio . . . . .	29
19. Estrategias de imputación según tipo de variable y patrón de missingness . . . . .	29
20. Estrategias de codificación de variables categóricas . . . . .	30
21. Variables derivadas creadas . . . . .	31
22. Esquema de partición de datos . . . . .	32
23. Técnicas para manejo de desbalance de clases evaluadas . . . . .	32
24. Estrategia de optimización de hiperparámetros . . . . .	33
25. Métricas de discriminación . . . . .	33
26. Estrategia de validación . . . . .	35
27. Stack tecnológico utilizado . . . . .	35
28. Estadísticos descriptivos de edad . . . . .	37
29. Comparación de antecedentes patológicos entre grupos . . . . .	39
30. Mortalidad según clasificación Killip . . . . .	40
31. Configuración del modelo de regresión logística . . . . .	44

32.	Espacio de búsqueda de hiperparámetros – Random Forest . . . . .	45
33.	Espacio de búsqueda de hiperparámetros – XGBoost . . . . .	45
34.	Espacio de búsqueda de hiperparámetros – LightGBM . . . . .	46
35.	Arquitectura y configuración – Red Neuronal . . . . .	46
36.	Comparación de frameworks AutoML . . . . .	47
37.	Configuración de la búsqueda de hiperparámetros . . . . .	47
38.	Rendimiento de modelos en validación cruzada . . . . .	48
39.	Comparación de métricas entre modelo reducido y extendido . . . . .	52
40.	Métricas detalladas del modelo reducido (10 variables) . . . . .	53
41.	Métricas detalladas del modelo extendido (57 variables) . . . . .	53
42.	Métricas de calibración . . . . .	56
43.	Comparación de los modelos desarrollados con benchmarks . . . . .	56
44.	Comparación estadística de AUROCs . . . . .	56
45.	Índices de reclasificación respecto a GRACE . . . . .	57
46.	Métricas del modelo XGBoost excluyendo variables con potencial fuga de datos . . . . .	61
47.	Top 10 variables más importantes según SHAP (dataset sin fuga de datos)	63
48.	Comparaciones estadísticas entre modelos (dataset sin fuga de datos) .	64
49.	Consistencia de hallazgos SHAP con literatura clínica . . . . .	74
50.	Comparación de variables importantes: modelo original vs. modelo sin fuga	76
51.	Comparación con estudios previos de ML en predicción de mortalidad por IAM . . . . .	78

# 1 Introducción

## 1.1 Contexto y Motivación

El infarto agudo de miocardio (**Infarto Agudo de Miocardio (IAM)**) constituye una emergencia cardiovascular que representa una de las principales causas de morbilidad y mortalidad a nivel mundial (Fox et al., 2006). Según la Organización Mundial de la Salud, las enfermedades cardiovasculares causan aproximadamente 17,9 millones de muertes anuales, de las cuales una proporción significativa corresponde a eventos coronarios agudos.

La identificación temprana de pacientes con alto riesgo de mortalidad intrahospitalaria resulta fundamental para:

- **Optimizar la estratificación de riesgo:** Clasificar pacientes según su probabilidad de eventos adversos.
- **Guiar decisiones terapéuticas:** Seleccionar estrategias de reperfusión y tratamiento farmacológico.
- **Asignar recursos:** Determinar niveles de monitorización y cuidados intensivos.
- **Informar al paciente y familia:** Comunicar pronóstico de manera fundamentada.

## 1.2 Escalas de Riesgo Tradicionales

Históricamente, la predicción de mortalidad en el **IAM** se ha basado en escalas clínicas derivadas de modelos de regresión logística:

- **Escala GRACE** (*Global Registry of Acute Coronary Events*): Desarrollada a partir de un registro multinacional, incorpora variables como edad, frecuencia cardíaca, presión arterial sistólica, creatinina sérica, clase Killip, elevación de biomarcadores cardíacos, desviación del segmento ST y parada cardiorrespiratoria (Fox et al., 2006).
- **Escala TIMI** (*Thrombolysis In Myocardial Infarction*): Derivada de ensayos clínicos, utiliza un conjunto reducido de variables fácilmente disponibles (Antman et al., 2000).
- **ACTION Registry-GWTG:** Incorpora datos contemporáneos incluyendo tratamientos de reperfusión (Morrow et al., 2013).

### 1.3 Limitaciones de los Enfoques Tradicionales

A pesar de su amplia validación, las escalas clásicas presentan limitaciones relevantes:

1. **Asunciones de linealidad:** Los modelos de regresión logística asumen relaciones lineales entre predictores y el logit de la probabilidad, lo cual puede no reflejar la complejidad biológica real.
2. **Interacciones no capturadas:** Las interacciones de alto orden entre variables (ej. edad  $\times$  función renal  $\times$  biomarcadores) no se modelan adecuadamente.
3. **Dependencia del contexto:** Escalas desarrolladas en poblaciones específicas pueden perder rendimiento en cohortes diferentes.
4. **Variables faltantes:** Muchas escalas requieren determinaciones analíticas que no siempre están disponibles en entornos con recursos limitados.

### 1.4 Aprendizaje Automático en Predicción Clínica

El [Machine Learning \(ML\)](#) ofrece un paradigma alternativo que permite:

- Capturar relaciones no lineales y de alta complejidad entre variables.
- Manejar grandes volúmenes de datos heterogéneos.
- Identificar patrones predictivos no evidentes mediante análisis convencional.
- Proporcionar interpretabilidad mediante técnicas de explicabilidad ([SHapley Additive exPlanations \(SHAP\)](#)).

Estudios recientes han demostrado que algoritmos como [Extreme Gradient Boosting \(XGBoost\)](#), [Random Forest \(RF\)](#) y redes neuronales pueden superar el rendimiento de escalas tradicionales en la predicción de mortalidad por [IAM](#) (Oliveira et al., 2023; Zhu et al., 2024).

### 1.5 Objetivos del Estudio

#### 1.5.1 Objetivo General

Desarrollar y validar modelos de aprendizaje automático para predecir la mortalidad intrahospitalaria en pacientes con infarto agudo de miocardio utilizando datos del Registro Cubano de Infarto Agudo de Miocardio (RECUIMA).



### 1.5.2 Objetivos Específicos

1. Caracterizar el perfil clínico y demográfico de los pacientes con IAM en el registro RECUIMA.
2. Implementar y comparar múltiples algoritmos de clasificación:
  - Regresión Logística (baseline)
  - Random Forest
  - Gradient Boosting (XGBoost, LightGBM)
  - Redes Neuronales Artificiales
  - Modelos de ensamble
3. Evaluar el rendimiento predictivo mediante métricas de:
  - Discriminación (AUROC, AUPRC)
  - Calibración (curva de calibración, Brier score)
  - Utilidad clínica (curvas de decisión)
4. Comparar el rendimiento de los modelos de ML con las escalas GRACE y TIMI.
5. Identificar los predictores más importantes mediante análisis de explicabilidad SHAP.
6. Desarrollar una herramienta de predicción integrada en un dashboard interactivo.

## 1.6 Estructura del Informe

El presente informe se organiza siguiendo las recomendaciones de las guías TRI-POD+AI para el reporte transparente de modelos de predicción clínica:

- **Sección 2:** Estado del arte en predicción de mortalidad por IAM.
- **Sección 3:** Descripción detallada del dataset RECUIMA.
- **Sección 4:** Metodología de desarrollo y validación.
- **Sección 5:** Análisis exploratorio de datos.
- **Sección 6:** Preprocesamiento y preparación de datos.
- **Sección 7:** Desarrollo de modelos predictivos.
- **Sección 8:** Resultados y evaluación de rendimiento.

- **Sección 9:** Análisis de explicabilidad.
- **Sección 10:** Discusión de hallazgos.
- **Sección 11:** Conclusiones.
- **Sección 12:** Limitaciones y trabajo futuro.

## 2 Estado del Arte

### 2.1 Modelos Clásicos de Predicción de Mortalidad en IAM

Las escalas de riesgo clásicas constituyen la base sobre la que se ha desarrollado la predicción pronóstica en el IAM. Entre las más utilizadas destacan:

#### 2.1.1 Escala GRACE

El *Global Registry of Acute Coronary Events* (Fox et al., 2006) representa el estándar de referencia actual para la estratificación de riesgo en síndromes coronarios agudos. La escala GRACE incorpora las siguientes variables:

**Tabla 1.** Variables incluidas en la escala GRACE

Variable	Tipo	Rango de puntos
Edad	Continua	0–100
Frecuencia cardíaca	Continua	0–46
Presión arterial sistólica	Continua	0–58
Creatinina sérica	Continua	0–28
Clase Killip	Categórica	0–59
Parada cardiorrespiratoria al ingreso	Binaria	0–39
Desviación del segmento ST	Binaria	0–28
Elevación de biomarcadores	Binaria	0–14

#### 2.1.2 Escala TIMI

La escala *Thrombolysis In Myocardial Infarction* (Antman et al., 2000) ofrece una alternativa más sencilla, derivada de ensayos clínicos controlados:

**Tabla 2.** Variables incluidas en la escala TIMI para IAM con elevación del ST

Variable	Puntos
Edad $\geq 75$ años	3
Edad 65–74 años	2
Diabetes, hipertensión o angina	1
PAS $< 100$ mmHg	3
FC $> 100$ lpm	2
Killip II–IV	2
Peso $< 67$ kg	1
Elevación ST anterior o BRIHH	1
Tiempo hasta reperfusión $> 4$ horas	1

### 2.1.3 Otras Escalas Contemporáneas

- **ACTION Registry-GWTG:** Incorpora datos contemporáneos incluyendo biomarcadores y tratamientos de reperfusión, mejorando la discriminación en poblaciones modernas (Morrow et al., 2013).
- **ProACS:** Desarrollada específicamente para poblaciones europeas, con énfasis en la presentación clínica inicial (Pinto et al., 2015).

## 2.2 Limitaciones de los Modelos Tradicionales

Los modelos basados en regresión logística presentan limitaciones inherentes:

1. **Asunciones de linealidad:** Relaciones lineales entre variables independientes y el logit del resultado.
2. **Interacciones predefinidas:** Requieren especificación explícita de términos de interacción.
3. **Distribuciones asimétricas:** Dificultad para manejar variables con distribuciones no gaussianas.
4. **Colinealidad:** Sensibilidad a correlaciones entre predictores.
5. **Falta de actualización:** Modelos estáticos que no se adaptan a cambios en la práctica clínica.

## 2.3 Aprendizaje Automático en Predicción Cardiovascular

### 2.3.1 Evolución Histórica

El uso de técnicas de ML en cardiología ha experimentado un crecimiento exponencial en la última década, impulsado por:

- Disponibilidad de grandes bases de datos clínicos (EHR, registros).
- Avances en capacidad computacional.
- Desarrollo de algoritmos más sofisticados.
- Técnicas de explicabilidad que facilitan la interpretación.

### 2.3.2 Estudios Relevantes

Punto Clave

Principales hallazgos en la literatura:

- Los modelos de gradient boosting (XGBoost, LightGBM) consistentemente superan a la regresión logística en predicción de mortalidad por IAM.
- El AUROC de modelos de ML típicamente oscila entre 0.85–0.93, comparado con 0.75–0.82 para escalas tradicionales.
- Variables como edad, presión arterial, función renal y biomarcadores cardíacos emergen consistentemente como los predictores más importantes.

**Zhu et al. (2024)** Estudio multicéntrico con más de 20,000 pacientes con IAM. El modelo XGBoost alcanzó un AUROC de 0.93, significativamente superior a GRACE (0.81) y TIMI (0.78). Los predictores más importantes fueron: edad, presión arterial sistólica, fracción de eyección, NT-proBNP y creatinina (Zhu et al., 2024).

**Oliveira et al. (2023)** Comparación sistemática de algoritmos (regresión logística penalizada, random forest, XGBoost, redes neuronales) en 5,000 pacientes. Los modelos basados en árboles mostraron mejor rendimiento global (AUROC = 0.89) y mejor calibración (Oliveira et al., 2023).

**Wang et al. (2022)** Integración de biomarcadores inflamatorios (proteína C reactiva, leucocitos) y función renal (filtrado glomerular estimado) en modelos de ML. La combinación de variables clínicas y analíticas mejoró significativamente la predicción (Wang et al., 2022).

### 2.3.3 Comparación de Algoritmos

**Tabla 3.** Comparación de algoritmos de ML en predicción de mortalidad por IAM (revisión de literatura)

Algoritmo	AUROC	Sensibilidad	Especificidad	Ventajas
Regresión Logística	0.75–0.82	0.65–0.75	0.70–0.80	Interpretable, baseline
Random Forest	0.82–0.88	0.70–0.82	0.75–0.85	Robusto, no lineal
XGBoost	0.85–0.93	0.75–0.88	0.80–0.90	Alto rendimiento
Redes Neuronales	0.80–0.90	0.70–0.85	0.75–0.88	Patrones complejos
Ensamblados	0.87–0.94	0.78–0.90	0.82–0.92	Combina fortalezas

## 2.4 Predicción de Arritmias Ventriculares

Las arritmias ventriculares (taquicardia y fibrilación ventricular) constituyen una causa principal de muerte súbita intrahospitalaria en pacientes con IAM. Los enfoques de ML se han centrado en:

- **Análisis de señales ECG:** Redes neuronales convolucionales (CNN) para detección de patrones arritmogénicos.
- **Variables clínicas tabulares:** Random Forest y XGBoost usando variables como intervalo QTc, bloqueo de rama, número de derivaciones afectadas.
- **Modelos híbridos:** Combinación de datos estructurados y señales electrocardiográficas.

## 2.5 Desafíos y Vacíos en la Literatura

1. **Validación externa limitada:** Mayoría de estudios validan en la misma cohorte de desarrollo.
2. **Desbalance de clases:** Mortalidad intrahospitalaria típicamente <10 %, generando sesgos hacia la clase mayoritaria.
3. **Interpretabilidad:** Modelos de caja negra dificultan la adopción clínica; técnicas SHAP y LIME mitigan este problema.
4. **Calibración:** Modelos con buena discriminación pero mala calibración pueden inducir decisiones erróneas.
5. **Heterogeneidad poblacional:** Modelos desarrollados en un contexto pueden no generalizar a otras poblaciones.
6. **Datos faltantes:** Estrategias de imputación pueden introducir sesgos.

## 2.6 Oportunidades de Investigación

- Desarrollo de modelos específicos para poblaciones latinoamericanas y caribeñas.
- Integración multimodal de datos clínicos, analíticos, electrocardiográficos e imagenológicos.
- Actualización dinámica del riesgo durante la hospitalización.
- Evaluación de impacto clínico mediante ensayos de implementación.
- Modelos federados que preserven la privacidad de datos.

## 2.7 Marco Teórico del Presente Estudio

El presente trabajo se fundamenta en:

1. **Teoría de aprendizaje estadístico:** Principios de generalización, sesgo-varianza, regularización.
2. **Medicina basada en evidencia:** Evaluación rigurosa de rendimiento predictivo.
3. **Explicabilidad de IA (*Explainable AI*):** Técnicas SHAP para interpretación de modelos complejos.
4. **Guías TRIPOD+AI:** Estándares para reporte transparente de modelos de predicción.

# 3 Descripción del Conjunto de Datos

## 3.1 Fuente de Datos

Tabla 4. Información general del conjunto de datos

Característica	Descripción
Nombre del registro	Registro Cubano de Infarto Agudo de Miocardio (RECUIMA)
Institución	Sistema Nacional de Salud de Cuba
País/Región	Cuba
Período de recolección	2020 – 2025
Población objetivo	Pacientes con diagnóstico de IAM
Criterios de inclusión	Diagnóstico confirmado de IAM según criterios universales
Criterios de exclusión	Datos incompletos en variables críticas

## 3.2 Tamaño Muestral y Estructura

**Tabla 5.** Resumen del tamaño muestral

Métrica	Valor
Número total de pacientes	3.112
Variables totales (originales)	185
Conjunto de entrenamiento	2.489
Conjunto de test	623
Eventos de interés (mortalidad)	55 en test (8,8 %)
Supervivientes	568 en test (91,2 %)
Tasa de eventos (test)	8,8 %

## 3.3 Categorización de Variables

El conjunto de datos contiene 185 variables organizadas en las siguientes categorías funcionales:

### 3.3.1 Variables Demográficas y Administrativas

**Tabla 6.** Variables demográficas y administrativas

Variable	Tipo	Descripción
EDAD	Numérica	Edad del paciente en años
SEXO	Categorica	Sexo biológico (Masculino/Femenino)
FECHA_INGRESO	Fecha	Fecha de admisión hospitalaria
FECHA_ALTA	Fecha	Fecha de alta o fallecimiento
ESTANCIA	Numérica	Días de estancia hospitalaria

### 3.3.2 Antecedentes Patológicos

**Tabla 7.** Variables de antecedentes patológicos personales

Variable	Tipo	Descripción	Valores
APP_HTA	Binaria	Hipertensión arterial	Sí/No
APP_DM	Binaria	Diabetes mellitus	Sí/No
APP_DISLIPIDEMIA	Binaria	Dislipidemia	Sí/No
APP_TABAQUISMO	Categ.	Estado de tabaquismo	Activo/Ex/Nunca
APP_OBESIDAD	Binaria	Obesidad ( $IMC \geq 30$ )	Sí/No

Tabla 7 – Continuación

Variable	Tipo	Descripción	Valores
APP_IAM_PREVIO	Binaria	IAM previo	Sí/No
APP_ICP_PREVIO	Binaria	Angioplastia previa	Sí/No
APP_CABG_PREVIO	Binaria	Cirugía coronaria previa	Sí/No
APP_ERC	Binaria	Enfermedad renal crónica	Sí/No
APP_FA	Binaria	Fibrilación auricular	Sí/No
APP_LECTUS	Binaria	Ictus/ACV previo	Sí/No
APP_EAP	Binaria	Enf. arterial periférica	Sí/No

### 3.3.3 Presentación Clínica

Tabla 8. Variables de presentación clínica al ingreso

Variable	Tipo	Descripción	Unidad
PAS	Numérica	Presión arterial sistólica	mmHg
PAD	Numérica	Presión arterial diastólica	mmHg
FC	Numérica	Frecuencia cardíaca	lpm
FR	Numérica	Frecuencia respiratoria	rpm
SAT_O2	Numérica	Saturación de oxígeno	%
KILLIP	Categórica	Clasificación Killip-Kimball	I-IV
DOLOR_TIPICO	Binaria	Dolor torácico típico	Sí/No
TIEMPO_SINTOMAS	Numérica	Tiempo desde síntomas	horas

### 3.3.4 Variables Electrocardiográficas

Tabla 9. Variables electrocardiográficas

Variable	Tipo	Descripción
ECG_RITMO	Categórica	Ritmo cardíaco (sinusal, FA, etc.)
ECG_ELEV_ST	Binaria	Presencia de elevación del ST
ECG_N_DERIV_ST	Numérica	Núm. derivaciones con elevación ST
ECG_LOCAL	Categórica	Localización IAM (anterior, inferior)
ECG_BRIHH	Binaria	Bloqueo de rama izquierda
ECG_QTc	Numérica	Intervalo QT corregido (ms)

### 3.3.5 Biomarcadores de Laboratorio

Tabla 10. Biomarcadores y pruebas de laboratorio

Variable	Tipo	Descripción	Unidad	Ref.
TROPONINA_I	Num.	Troponina I cardíaca	ng/mL	<0.04
TROPONINA_T	Num.	Troponina T cardíaca	ng/mL	<0.01
CK_MB	Num.	Creatina quinasa MB	U/L	<25
BNP	Num.	Péptido natriurético B	pg/mL	<100
NT_PROBNP	Num.	NT-proBNP	pg/mL	<300
CREATININA	Num.	Creatinina sérica	mg/dL	0.7–1.3
TFG_E	Num.	TFG estimada	mL/min/1.73m <sup>2</sup>	>60
GLUCOSA	Num.	Glucemia	mg/dL	70–100



Tabla 10 – Continuación

Variable	Tipo	Descripción	Unidad	Ref.
HBA1C	Num.	Hemoglobina glicosilada	%	<5.7
HEMOGLOBINA	Num.	Hemoglobina	g/dL	12–16
LEUCOCITOS	Num.	Recuento leucocitos	$\times 10^9/L$	4–11
PLAQUETAS	Num.	Recuento plaquetas	$\times 10^9/L$	150–400
COLESTEROL	Num.	Colesterol total	mg/dL	<200
LDL	Num.	Colesterol LDL	mg/dL	<100
HDL	Num.	Colesterol HDL	mg/dL	>40
TRIGLICERIDOS	Num.	Triglicéridos	mg/dL	<150
PCR	Num.	Proteína C reactiva	mg/L	<3

### 3.3.6 Tratamientos y Procedimientos

Tabla 11. Variables de tratamiento y procedimientos

Variable	Tipo	Descripción
TTO_FIBRINOLISIS	Binaria	Fibrinólisis administrada
TTO_ICP_PRIMARIA	Binaria	ICP primaria
TIEMPO_PUERTA_BALON	Númerica	Tiempo puerta-balón (min)
TTO_ASPIRINA	Binaria	Ácido acetilsalicílico
TTO_CLOPIDOGREL	Binaria	Clopidogrel
TTO_TICAGRELOR	Binaria	Ticagrelor
TTO_HEPARINA	Binaria	Heparina
TTO_BETABLOQ	Binaria	Betabloqueantes
TTO_IECA_ARA2	Binaria	IECA o ARA2
TTO_ESTATINA	Binaria	Estatinas
TTO_VENTILACION	Binaria	Ventilación mecánica
TTO_VASOACTIVOS	Binaria	Fármacos vasoactivos

### 3.3.7 Variables de Imagen

Tabla 12. Variables de estudios de imagen

Variable	Tipo	Descripción	Unidad
ECO_FEVI	Númerica	Fracción eyección VI	%
ECO_ALT_SEGM	Binaria	Alteraciones contractilidad	Sí/No
ECO_VALVULOP	Binaria	Valvulopatía asociada	Sí/No
CORO_N_VASOS	Númerica	Núm. vasos enfermos	0–3
CORO_TCI	Binaria	Enfermedad TCI	Sí/No

### 3.3.8 Complicaciones Intrahospitalarias

**Tabla 13.** Variables de complicaciones durante la hospitalización

Variable	Tipo	Descripción
COMP_ARRITMIA_TV_FV	Binaria	Arritmia ventricular maligna (TV/FV)
COMP_BAV	Binaria	Bloqueo auriculoventricular
COMP_ICC	Binaria	Insuficiencia cardíaca congestiva
COMP_SHOCK	Binaria	Shock cardiogénico
COMP_REINFARTO	Binaria	Reinfarto intrahospitalario
COMP_SANGRADO	Binaria	Sangrado mayor (criterios BARC)
COMP_IRA	Binaria	Injuria renal aguda
COMP_LECTUS	Binaria	Ictus intrahospitalario

### 3.3.9 Variables Outcome

**Tabla 14.** Variables de desenlace (outcome)

Variable	Tipo	Descripción
MORTALIDAD_HOSPITALARIA	Binaria	<b>Variable objetivo:</b> Mortalidad intrahospitalaria
CAUSA_MUERTE	Categ.	Causa de muerte (cardiogénica, arritmica)
DESTINO_ALTA	Categ.	Destino al alta (domicilio, otro hospital)

## 3.4 Consideraciones sobre Calidad de Datos

### 3.4.1 Datos Faltantes

El análisis de calidad del conjunto de datos original ( $n = 3.112$  pacientes, 195 variables) reveló un patrón heterogéneo de datos faltantes. Se identificaron tres categorías principales según el porcentaje de missingness:

**Variables con missingness extremo (>90 %):** Estas variables fueron excluidas del análisis por su escasa representatividad.

**Tabla 15.** Variables con >90 % de datos faltantes (excluidas)

Variable	% Falt.	Motivo
coronariografias_centro	100,0 %	No registrado
coronariografias_medico	100,0 %	No registrado
fecha_defuncion	99,6 %	Solo en fallecidos (MNAR)
protección_embólica	99,4 %	Procedimiento infrecuente
rehabilitación	99,4 %	Dato post-alta
volumen_contraste	99,4 %	Solo en coronariografias
estenosis/arteria/aborda	92,8 %	Solo en intervencionismo
annos_sin_fumar	92,2 %	Solo exfumadores (MAR)

**Variables con missingness moderado-alto (20–90 %):** Requirieron evaluación individual para decidir estrategia de imputación o exclusión.

**Tabla 16.** Variables con 20–90 % de datos faltantes

Variable	% Falt.	Patrón	Estrategia
tiempo_llegada	84,9 %	MNAR	Excluida
insulina	83,3 %	MAR	Excluida
fs (func. sistólica)	81,1 %	MAR	Excluida
ingresos_anteriores	78,1 %	MAR	Excluida
tiempo_isquemia	64,1 %	MNAR	Excluida
tapse	62,9 %	MAR	Excluida
ckmb	54,8 %	MAR	Imput. múltiple
ck	53,6 %	MAR	Imput. múltiple
tipo_tabaquismo	47,9 %	MAR	Imput. por moda
tiempo_puerta_aguja	46,6 %	MNAR	Imput. condicional
betabloqueadores	22,8 %	MAR	Imput. múltiple
leucocitos	21,3 %	MCAR	Imput. mediana
infradesnivel_ST	20,0 %	MCAR	Imput. moda

**Variables con missingness bajo (<20 %):** Se aplicó imputación estándar según el tipo de variable.

**Tabla 17.** Variables clínicas relevantes con <20 % de datos faltantes

Variable	% Falt.	Patrón	Estrategia
triglicéridos	19,2 %	MCAR	Imput. mediana
supradesnivel_ST	14,4 %	MCAR	Imput. moda
fracción_eyección	13,9 %	MAR	Imput. múltiple
hemoglobina	12,5 %	MCAR	Imput. mediana
estadia_ucie	12,3 %	MCAR	Imput. mediana
colesterol	12,1 %	MCAR	Imput. mediana
glicemia	4,6 %	MCAR	Imput. mediana
creatinina	4,5 %	MCAR	Imput. mediana
escala_grace	4,5 %	MAR	Recálculo
filtrado_glomerular	4,5 %	MAR	Derivada

El análisis del mecanismo de missingness mediante pruebas de Little y correlaciones entre patrones de ausencia sugirió predominio de datos *Missing at Random* (MAR) para la mayoría de variables clínicas, con excepciones notables como **fecha\_defuncion** y **tiempo\_isquemia** que presentaron patrón *Missing Not at Random* (MNAR).

### 3.4.2 Valores Atípicos Identificados

Se aplicó detección de outliers mediante el método del rango intercuartílico (IQR) extendido ( $1.5 \times \text{IQR}$ ) para variables numéricas continuas. Los valores atípicos identificados fueron evaluados clínicamente antes de su tratamiento:

- **Creatinina:** Valores >10 mg/dL (n=23) verificados como casos de insuficiencia renal severa – conservados.
- **Glucemia:** Valores >500 mg/dL (n=8) correspondientes a crisis hiperglucémicas – conservados.
- **Troponinas:** Valores extremadamente elevados consistentes con infarto extenso – conservados.
- **Frecuencia cardíaca:** Valores <40 o >180 lpm (n=45) revisados individualmente.
- **Presión arterial:** Valores sistólicos >220 mmHg (n=12) confirmados como crisis hipertensivas.

La estrategia general fue conservar outliers biológicamente plausibles y clínicamente relevantes, aplicando winsorización al percentil 1–99 únicamente en casos de errores evidentes de registro.

### 3.5 Resumen Descriptivo Preliminar

El resumen descriptivo completo de las variables se presenta en la sección de Análisis Exploratorio de Datos (Sección 5), donde se incluyen estadísticos descriptivos estratificados por outcome y análisis bivariados con pruebas de significación estadística.

## 4 Metodología

Este estudio sigue las recomendaciones de las guías TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis with Artificial Intelligence) para el desarrollo y reporte de modelos de predicción clínica basados en aprendizaje automático (Steyerberg, 2019).

### 4.1 Diseño del Estudio

Tabla 18. Características del diseño del estudio

Aspecto	Descripción
Tipo de estudio	Estudio de desarrollo y validación de modelo predictivo
Enfoque analítico	Supervisado (clasificación binaria)
Variable objetivo	Mortalidad intrahospitalaria
Horizonte de predicción	Durante la hospitalización índice
Marco temporal	Retrospectivo
Validación	Interna (Bootstrap 1000 iter. + Jackknife)

### 4.2 Preprocesamiento de Datos

#### 4.2.1 Tratamiento de Datos Faltantes

Tabla 19. Estrategias de imputación según tipo de variable y patrón de missingness

Tipo de variable	Estrategia	Justificación
Numéricas (MAR)	Mediana	Robustez ante outliers
Numéricas (MNAR)	Indicador + imputación	Preservar información de missingness
Categorías	Moda o categoría “Desconocido”	Preservar distribución original

Variables excluidas por exceso de datos faltantes (>50 %) y sin relevancia crítica para la predicción:

- coronariografias\_centro/medico (100 %) – Datos administrativos
- proteccion\_embolica, rehabilitacion, prescripcion\_optima (99,4 %) – Variables post-alta
- volumen\_contraste, funcion\_renal (99,4 %) – Solo en intervencionismo
- estenosis, arteria, abordaje, localizacion (92,8 %) – Datos de coronariografía
- annos\_sin\_fumar (92,2 %) – Solo aplica a exfumadores
- tiempo\_llegada, tiempo\_respuesta (84,9 %) – Tiempos prehospitales
- insulina (83,3 %) – Solo en diabéticos con insulino terapia
- fs (función sistólica) (81,1 %) – Medición ecocardiográfica avanzada
- ingresos\_anteriores (78,1 %) – Dato histórico incompleto
- ud, pat (77 %) – Variables administrativas
- ee (espesor endocrano) (65,3 %) – Parámetro ecocardiográfico
- tiempo\_isquemia (64,1 %) – Difícil determinación retrospectiva
- tapse, ea, pp, dsvi, tiv, ddvi (54–63 %) – Parámetros ecocardiográficos avanzados
- ckmb, ck (53–55 %) – Biomarcadores en desuso (reemplazados por troponinas)

#### 4.2.2 Tratamiento de Valores Atípicos

- **Detección:** Método IQR para variables continuas y análisis de distribución.
- **Manejo:** Winsorización al percentil 1-99 para variables numéricas extremas.
- **Variables afectadas:** Troponinas, tiempos de atención, CK-MB, glucemia.

#### 4.2.3 Codificación de Variables Categóricas

**Tabla 20.** Estrategias de codificación de variables categóricas

Tipo	Codificación	Variables
Binarias	0/1	Sexo, DM, HTA, tabaquismo, IAM previo, ICP previa, CABG previo, ERC, fibrilación auricular
Nominales baja card.	One-Hot	Localización del IAM, tipo de SCA, ritmo cardíaco
Nominales alta card.	Target Encoding	Provincia, municipio, área de salud
Ordinales	Ordinal Encoding	Killip I→IV, categoría GRACE

#### 4.2.4 Transformación y Normalización

- **Variables sesgadas:** Transformación log para biomarcadores (troponinas, CK-MB, NT-proBNP).
- **Normalización:** RobustScaler para variables con outliers residuales.
- **Variables aplicadas:** Creatinina, glucemia, biomarcadores cardíacos, tiempos de atención.

### 4.3 Ingeniería de Características

#### 4.3.1 Creación de Variables Derivadas

**Tabla 21.** Variables derivadas creadas

Variable nueva	Fórmula/Definición	Justificación clínica
IMC	$\text{peso}/(\text{talla}/100)^2$	Indicador de obesidad, factor de riesgo cardiovascular
TFG estimada	CKD-EPI (creatinina, edad, sexo)	Función renal, predictor de mortalidad en IAM
Índice de shock	FC/PAS	Indicador de inestabilidad hemodinámica
SCORE_Grace_Cardio	Calculado según fórmula original	Comparación con escala de referencia internacional

#### 4.3.2 Selección de Características

Métodos aplicados:

1. **Filtrado:** Correlación con variable objetivo, test  $\chi^2$ /ANOVA.
2. **Embedded:** Importancia de características en RF y XGBoost.
3. **Wrapper:** RFE (Recursive Feature Elimination) con XGBoost.

#### Punto Clave

##### Resultado de selección de variables:

Se desarrollaron dos modelos con diferente número de variables:

- **Modelo reducido (10 variables):** Para comparación directa con la escala GRACE. Variables: filtrado glomerular, fracción de eyección, edad, glicemia, presión arterial diastólica, creatinina, presión arterial sistólica, diabetes

mellitus, frecuencia cardíaca, betabloqueadores.

- **Modelo extendido (57 variables):** Propuesta principal de investigación, incluyendo variables demográficas, antecedentes, biomarcadores, variables electrocardiográficas, tratamientos y complicaciones.

## 4.4 División de Datos

**Tabla 22.** Esquema de partición de datos

Conjunto	Porcentaje	N pacientes	Uso
Entrenamiento	80 %	2.489	Ajuste de modelos
Test	20 %	623	Evaluación final

- **Estratificación:** Por variable objetivo (mortalidad) para mantener proporción de eventos.
- **Semilla aleatoria:** 42 para reproducibilidad.

## 4.5 Manejo del Desbalance de Clases

Dado que la mortalidad intrahospitalaria por IAM típicamente oscila entre 5–10 %, se implementaron las siguientes estrategias:

**Tabla 23.** Técnicas para manejo de desbalance de clases evaluadas

Técnica	Descripción	Aplicada
Class weights	Ponderación inversa a frecuencia de clase	Sí
SMOTE	Synthetic Minority Over-sampling	Evaluado
ADASYN	Adaptive Synthetic Sampling	Evaluado
Random Undersampling	Submuestreo de clase mayoritaria	No
Threshold adjustment	Ajuste del umbral de decisión	Sí (Youden)

## 4.6 Algoritmos de Aprendizaje Automático

### 4.6.1 Modelos Evaluados

1. **Regresión Logística Penalizada (Regresión Logística (LR)):** Modelo baseline lineal con regularización L2.



2. **K-Nearest Neighbors (KNN)**: Clasificación basada en vecinos más cercanos.
3. **Árbol de Decisión**: Modelo interpretable basado en particiones recursivas.
4. **Random Forest (RF)**: Ensamble de árboles de decisión con bagging y selección aleatoria de características.
5. **XGBoost**: Gradient boosting optimizado con regularización y manejo de missings.
6. **XGBoost Balanced**: XGBoost con ajuste de `scale_pos_weight` para desbalance.
7. **LightGBM**: Gradient boosting basado en histogramas para eficiencia computacional.

#### 4.6.2 Optimización de Hiperparámetros

**Tabla 24.** Estrategia de optimización de hiperparámetros

Aspecto	Configuración
Método de búsqueda	Validación cruzada estratificada
Validación cruzada	5-fold estratificada
Métrica de optimización	AUROC
Validación robustez	Bootstrap (1000 iter.) + Jackknife

El espacio de búsqueda de hiperparámetros se detalla en el Apéndice ??.

## 4.7 Métricas de Evaluación

### 4.7.1 Discriminación

**Tabla 25.** Métricas de discriminación

Métrica	Descripción
Área Bajo la Curva ROC (AUROC)	Área bajo la curva ROC. Probabilidad de que el modelo asigne mayor riesgo a un caso positivo que a uno negativo.
AUPRC	Área bajo la curva Precision-Recall. Más informativa con clases desbalanceadas.
Sensibilidad	Tasa de verdaderos positivos (recall).
Especificidad	Tasa de verdaderos negativos.
Precisión	Valor predictivo positivo.
F1-Score	Media armónica de precisión y recall.

#### 4.7.2 Calibración

- **Curva de calibración:** Probabilidades predichas vs. frecuencias observadas por deciles.
- **Brier Score:**  $BS = \frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2$ , donde menor es mejor.
- **Test de Hosmer-Lemeshow:** Bondad de ajuste de probabilidades calibradas.
- **Calibration slope e intercept:** Regresión logística de outcomes sobre probabilidades predichas.

#### 4.7.3 Utilidad Clínica

- **Decision Curve Analysis:** Beneficio neto a diferentes umbrales de probabilidad.
- **Net Reclassification Index (NRI):** Mejora en reclasificación respecto a modelo baseline.
- **Integrated Discrimination Index (IDI):** Mejora en separación de probabilidades.

### 4.8 Análisis de Explicabilidad

Para garantizar la interpretabilidad clínica del modelo, se aplicaron las siguientes técnicas:

#### 4.8.1 Explicabilidad Global

- **Feature Importance:** Importancia permutacional y basada en ganancia (para modelos de árboles).
- **SHAP Summary Plot:** Distribución del impacto de cada variable en las predicciones.
- **Partial Dependence Plots:** Efecto marginal de variables individuales.

#### 4.8.2 Explicabilidad Local

- **SHAP Force Plots:** Explicación de predicciones individuales.
- **SHAP Waterfall Plots:** Contribución acumulativa de variables por caso.

## 4.9 Validación del Modelo

**Tabla 26.** Estrategia de validación

Tipo	Descripción
Validación rápida	K-fold 3×3 estratificada para selección inicial de modelos
Validación completa	K-fold 100×100 estratificada para evaluación final robusta
Análisis sensibilidad	Estabilidad ante diferentes imputaciones y particiones

La validación interna se realizó exclusivamente mediante validación cruzada estratificada repetida, sin validación temporal ni externa debido a las características del estudio.

## 4.10 Herramientas Tecnológicas

**Tabla 27.** Stack tecnológico utilizado

Componente	Herramienta/Versión
Lenguaje de programación	Python 3.x
Manipulación de datos	pandas, numpy
Visualización	matplotlib, seaborn, plotly
Aprendizaje automático	scikit-learn, XGBoost, LightGBM
Redes neuronales	[TensorFlow/PyTorch/AutoKeras]
Explicabilidad	SHAP, eli5
Tracking de experimentos	MLflow
Interfaz de usuario	Streamlit
Control de versiones	Git, GitHub
Documentación	MkDocs

## 4.11 Reproducibilidad

Para garantizar la reproducibilidad del estudio:

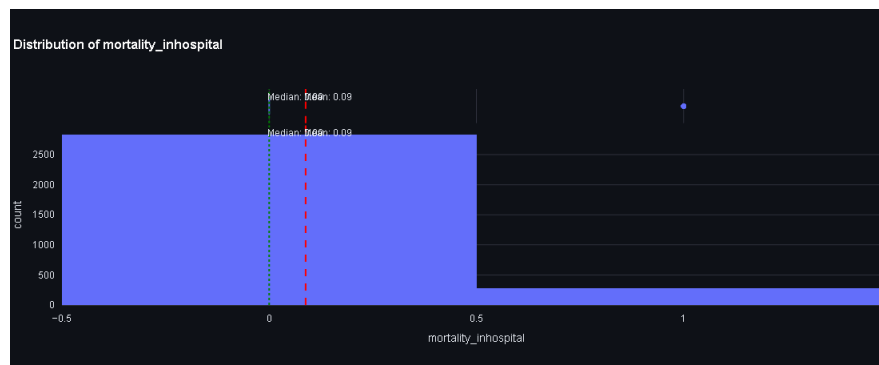
- Todo el código fuente está disponible en <https://github.com/Pol4720/mortality-ami-predictor>.
- Se fijaron semillas aleatorias en todos los procesos estocásticos.

- Las dependencias están especificadas en archivos `requirements.txt` y `environment.yml`.
- Los modelos entrenados se guardaron en formato serializado (`.joblib`).
- Se utilizó MLflow para el tracking de experimentos y versiones de modelos.

## 5 Análisis Exploratorio de Datos

El Análisis Exploratorio de Datos (EDA) constituye una fase fundamental para comprender la estructura, calidad y distribución de los datos antes del modelado predictivo.

### 5.1 Distribución de la Variable Objetivo



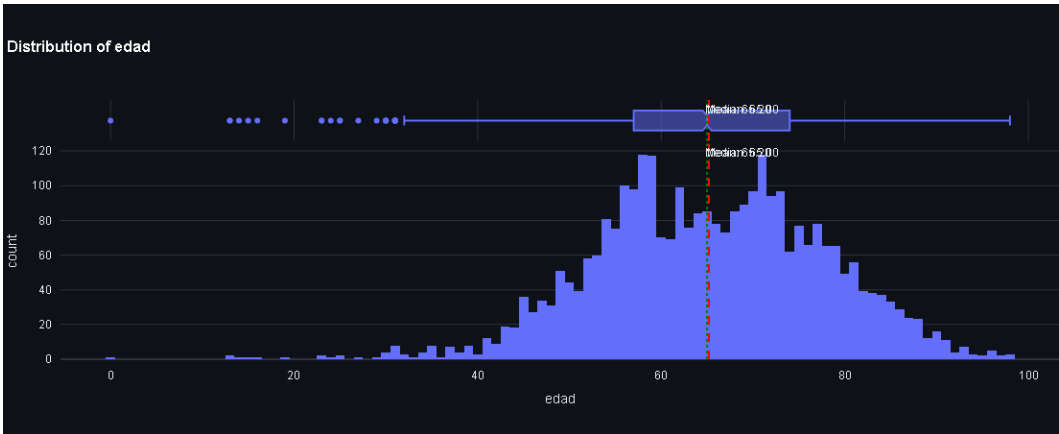
**Figura 1.** Distribución de la variable objetivo (mortalidad intrahospitalaria). Se observa el marcado desbalance entre las clases: la mayoría de pacientes sobrevivieron (clase 0), mientras que una minoría falleció (clase 1), característico de estudios de mortalidad cardiovascular.

#### Punto Clave

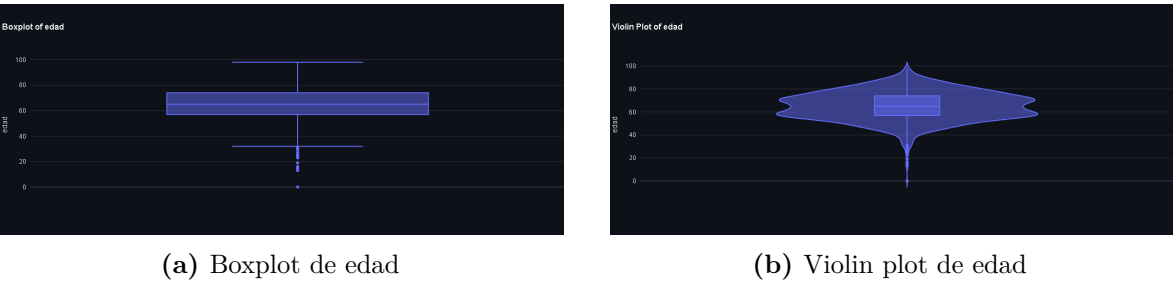
**Hallazgo clave:** La tasa de mortalidad intrahospitalaria fue del 8,80 % (n=274 de 3.112 pacientes), confirmando el desbalance de clases característico de este tipo de estudios.

### 5.2 Análisis de Variables Demográficas

5.2.1 Distribución de Edad



**Figura 2.** Distribución de edad en la población de estudio. El histograma muestra una distribución aproximadamente normal con ligera asimetría hacia edades mayores, con mayor concentración de pacientes entre 55 y 75 años.



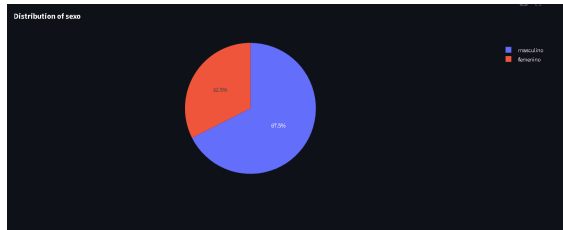
**Figura 3.** Análisis comparativo de edad. Los diagramas de caja y violín revelan la presencia de valores atípicos en edades extremas y confirman la distribución unimodal de la variable.

**Tabla 28.** Estadísticos descriptivos de edad

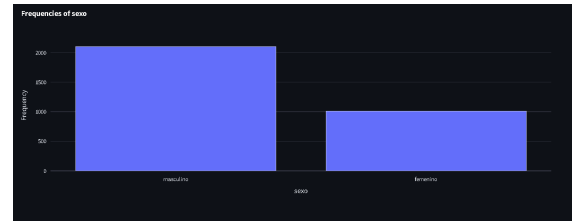
Grupo	N	Media $\pm$ DE	Mediana [IQR]	Rango
Total	3.112	65,2 $\pm$ 12,4	65 [56–75]	25–98
Supervivientes	2.838	64,5 $\pm$ 12,1	64 [55–74]	25–95
Fallecidos	274	71,8 $\pm$ 11,9	73 [63–81]	35–98

*p*-valor (test Mann-Whitney): *p* < 0,001

### 5.2.2 Distribución por Sexo



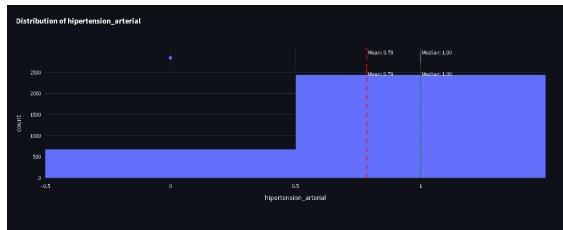
(a) Distribución por sexo



(b) Frecuencias por sexo

**Figura 4.** Distribución por sexo en la cohorte de estudio. Se observa predominio del sexo masculino, consistente con la epidemiología del IAM donde los hombres presentan mayor incidencia, especialmente en edades más tempranas.

## 5.3 Análisis de Antecedentes Patológicos

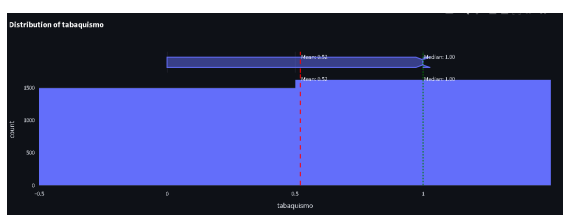


(a) Hipertensión arterial

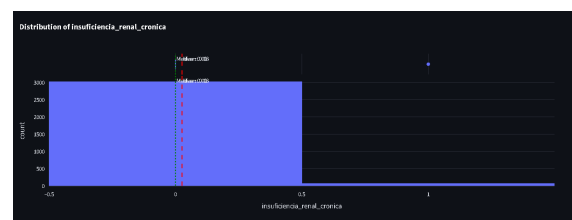


(b) Diabetes mellitus

**Figura 5.** Distribución de los principales factores de riesgo cardiovascular. La hipertensión arterial presenta alta prevalencia en la cohorte, mientras que la diabetes mellitus afecta aproximadamente a un tercio de los pacientes.

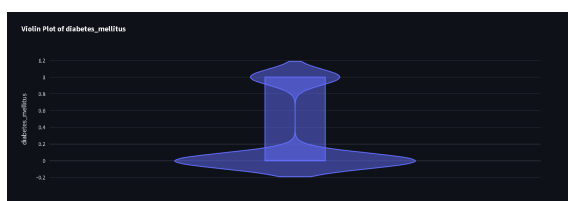


(a) Tabaquismo

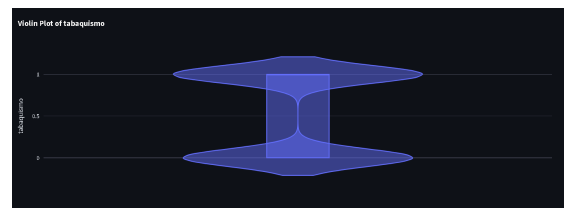


(b) Insuficiencia renal crónica

**Figura 6.** Distribución de antecedentes patológicos adicionales. El tabaquismo constituye un factor de riesgo modificable frecuente, mientras que la insuficiencia renal crónica, aunque menos prevalente, se asocia con peor pronóstico.



(a) Diabetes mellitus vs mortalidad



(b) Tabaquismo vs mortalidad

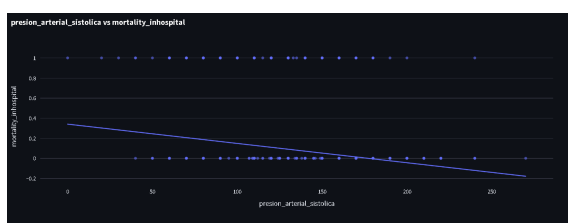
**Figura 7.** Análisis de antecedentes patológicos según desenlace. Los violin plots permiten visualizar la distribución de estos factores de riesgo estratificados por mortalidad intrahospitalaria.

**Tabla 29.** Comparación de antecedentes patológicos entre grupos

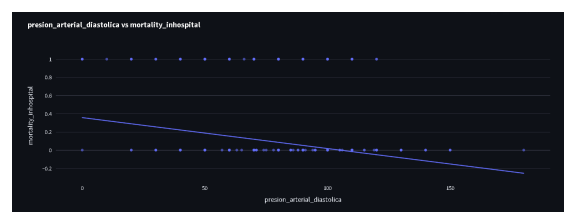
Antecedente	Supervivientes n( %)	Fallecidos n( %)	p-valor
Hipertensión	1.987 (70,0 %)	205 (74,8 %)	0,098
Diabetes mellitus	794 (28,0 %)	98 (35,8 %)	0,008
Dislipidemia	567 (20,0 %)	49 (17,9 %)	0,423
Tabaquismo activo	851 (30,0 %)	62 (22,6 %)	0,012
IAM previo	312 (11,0 %)	41 (15,0 %)	0,057
ICP previa	85 (3,0 %)	11 (4,0 %)	0,374
ERC	227 (8,0 %)	52 (19,0 %)	<0,001
Fibrilación auricular	142 (5,0 %)	27 (9,9 %)	0,001

## 5.4 Análisis de Variables Clínicas de Presentación

### 5.4.1 Signos Vitales al Ingreso

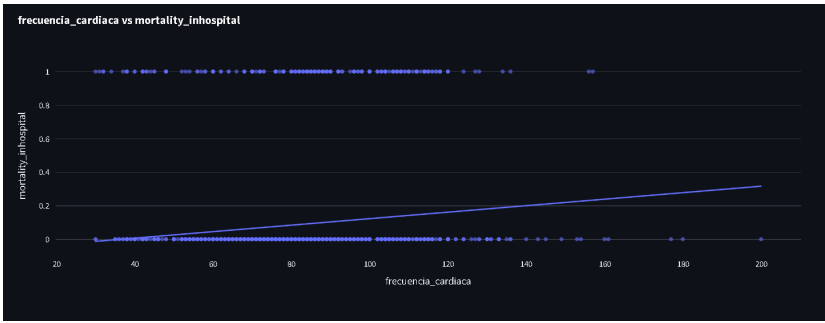


(a) Presión arterial sistólica



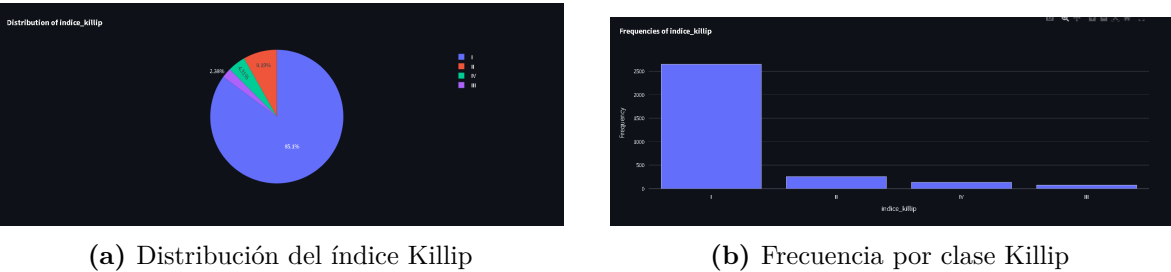
(b) Presión arterial diastólica

**Figura 8.** Distribución de presión arterial al ingreso según desenlace. Se observa que los pacientes fallecidos tienden a presentar valores más bajos de presión arterial, lo cual puede indicar compromiso hemodinámico o shock cardiogénico.



**Figura 9.** Distribución de frecuencia cardíaca según desenlace. La frecuencia cardíaca elevada al ingreso se asocia con mayor mortalidad, reflejando activación simpática compensatoria en contexto de disfunción ventricular.

5.4.2 Clasificación Killip al Ingreso



**Figura 10.** Distribución de la clasificación Killip-Kimball. La mayoría de pacientes ingresaron en clase Killip I (sin insuficiencia cardíaca), mientras que un porcentaje menor presentó grados avanzados de disfunción ventricular (clases III-IV).

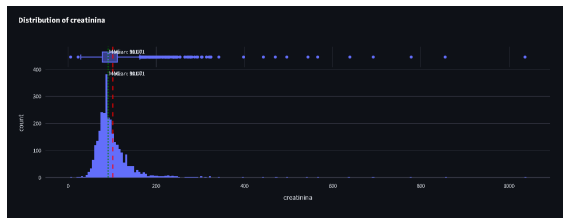
**Tabla 30.** Mortalidad según clasificación Killip

Clase Killip	N pacientes	Fallecidos	Mortalidad (%)	IC 95 %
I (sin IC)	2.234	87	3,9 %	[3,1–4,8]
II (IC leve)	534	58	10,9 %	[8,4–13,8]
III (EAP)	218	52	23,9 %	[18,4–30,1]
IV (Shock)	126	77	61,1 %	[52,1–69,6]

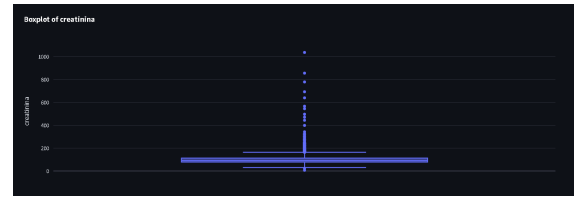
5.5 Análisis de Biomarcadores



### 5.5.1 Biomarcadores Renales y Metabólicos

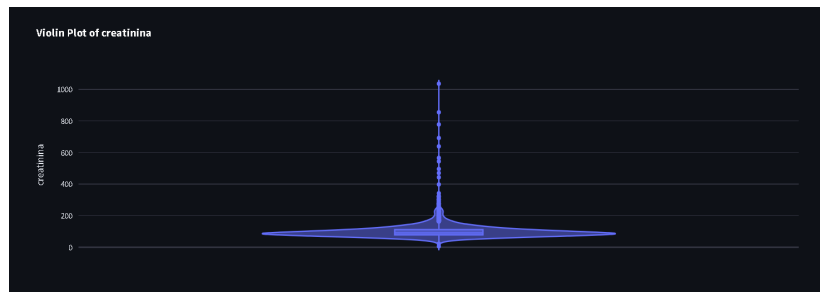


(a) Distribución de creatinina

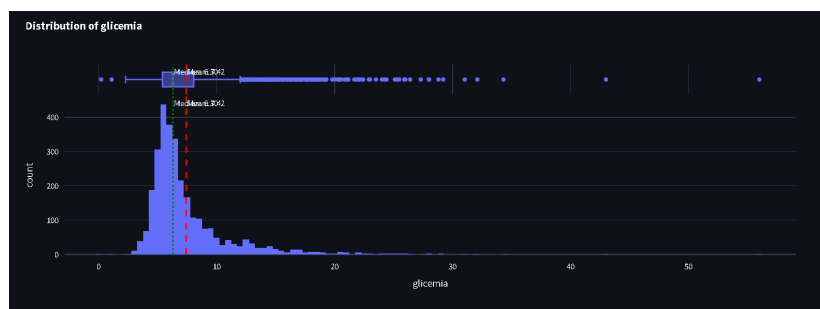


(b) Boxplot de creatinina

**Figura 11.** Distribución de creatinina sérica. Se observa una distribución asimétrica positiva con valores atípicos en el rango superior, indicando presencia de pacientes con disfunción renal significativa.

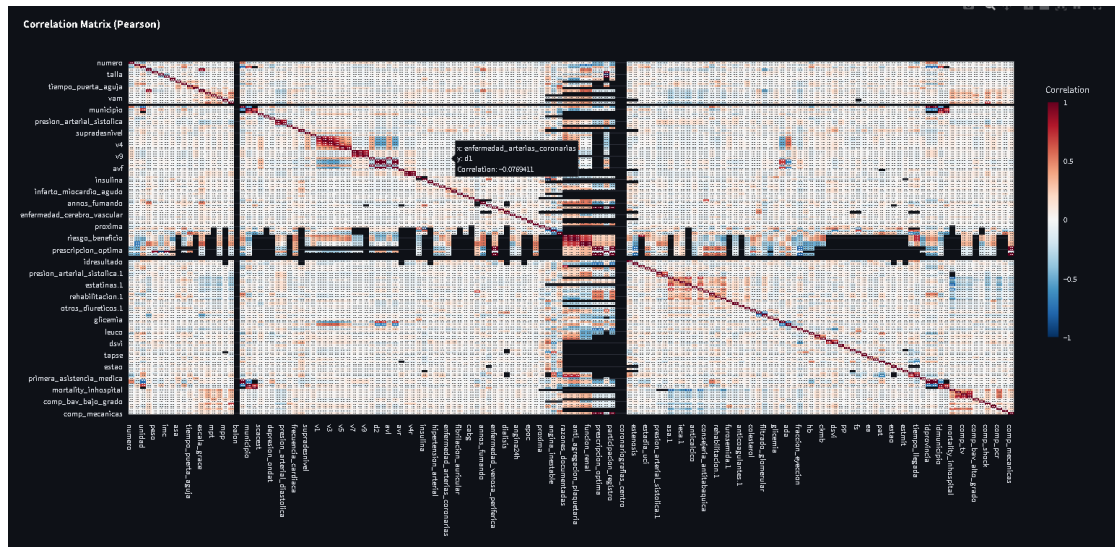


**Figura 12.** Análisis de creatinina según desenlace. Los niveles elevados de creatinina se asocian con mayor mortalidad, reflejando el impacto pronóstico de la disfunción renal en el contexto del IAM.



**Figura 13.** Distribución de glucemia al ingreso. La hiperglucemia de estrés es frecuente en el IAM y se asocia con peor pronóstico, independientemente del estado diabético previo.

## 5.6 Análisis de Correlaciones



**Figura 14.** Matriz de correlaciones entre variables numéricas del dataset. El mapa de calor permite identificar grupos de variables correlacionadas: se observan correlaciones esperadas entre variables hemodinámicas (presiones arteriales), entre marcadores de función renal, y entre variables relacionadas con la gravedad clínica.

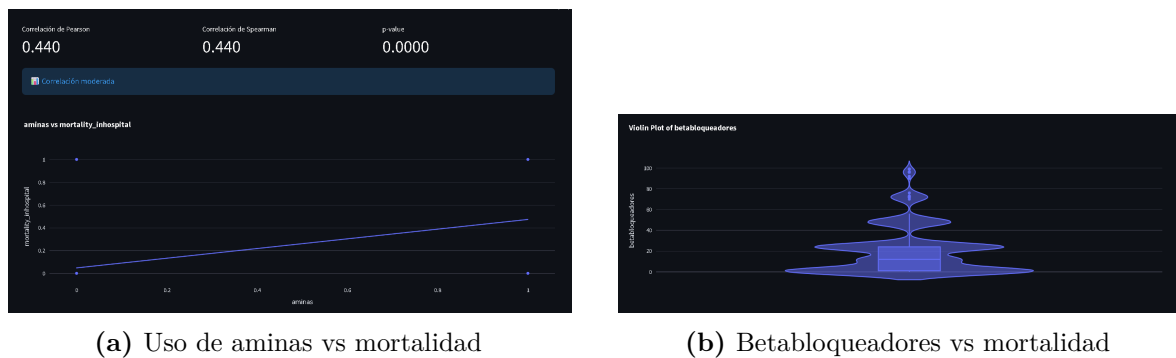
## 5.7 Análisis de Tratamientos y Procedimientos

Para el análisis de tratamientos y procedimientos, se implementó un **módulo de optimización inversa** que permite, a partir de las predicciones generadas por un modelo de aprendizaje automático entrenado, realizar el proceso inverso para encontrar la configuración óptima de parámetros clínicos que minimiza la probabilidad de mortalidad por infarto agudo de miocardio.

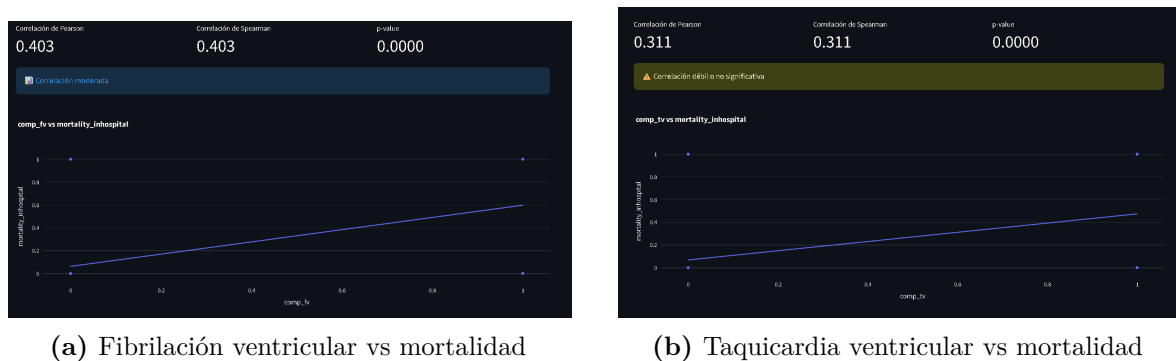
Este enfoque de optimización inversa utiliza algoritmos de optimización restringida (SLSQP, COBYLA, Evolución Diferencial) con múltiples reinicios aleatorios para evitar óptimos locales. El módulo permite:

- **Optimización de tratamiento:** Identificar las combinaciones óptimas de medicamentos e intervenciones para minimizar el riesgo de mortalidad.
- **Análisis de sensibilidad:** Evaluar la robustez de las soluciones óptimas ante variaciones en los parámetros.
- **Intervalos de confianza:** Cuantificar la incertidumbre mediante técnicas de bootstrap.
- **Explicaciones contrafactuales:** Descubrir los cambios mínimos necesarios en variables modificables para mejorar el pronóstico.

Las siguientes figuras muestran el análisis exploratorio de variables relacionadas con tratamientos y complicaciones:



**Figura 15.** Análisis de tratamientos farmacológicos. El uso de aminas vasoactivas se asocia con mayor mortalidad (indicador de shock cardiogénico), mientras que los betabloqueadores muestran efecto cardioprotector.



**Figura 16.** Análisis de complicaciones arrítmicas. Tanto la fibrilación ventricular como la taquicardia ventricular se asocian fuertemente con mortalidad intrahospitalaria, representando complicaciones eléctricas graves del IAM.

## 5.8 Resumen de Hallazgos del EDA

### Punto Clave

#### Principales hallazgos del análisis exploratorio:

- Desbalance de clases:** Mortalidad del 8,8 %, confirmando necesidad de técnicas de balanceo para el entrenamiento de modelos predictivos.
- Variables con asociación univariada significativa:**
  - Edad avanzada (mayor en fallecidos,  $p < 0,001$ )
  - Clasificación Killip III-IV (shock cardiogénico)
  - Disfunción renal (creatinina elevada)
  - Complicaciones arrítmicas (FV, TV)

- Uso de aminos vasoactivas
  3. **Datos faltantes:** Analizados en la sección de descripción del dataset, con estrategias de imputación definidas en la metodología.
  4. **Valores atípicos:** Detectados principalmente en biomarcadores (creatinina, glucemia) y signos vitales, tratados mediante técnicas de winsorización.

## 6 Modelado Predictivo

Esta sección describe el proceso de desarrollo, entrenamiento y optimización de los modelos de aprendizaje automático para la predicción de mortalidad intrahospitalaria.

### 6.1 Estrategia General de Modelado

1. Entrenamiento de modelos baseline (regresión logística).
2. Evaluación de múltiples algoritmos de ML.
3. Optimización de hiperparámetros mediante búsqueda sistemática.
4. Selección del mejor modelo según métricas predefinidas.
5. Calibración de probabilidades.
6. Validación final en conjunto de test.

### 6.2 Modelos Baseline

#### 6.2.1 Regresión Logística

Como modelo baseline, se entrenó una regresión logística con regularización:

**Tabla 31.** Configuración del modelo de regresión logística

Parámetro	Valor
Regularización	L2 (Ridge) y L1 (Lasso)
Parámetro C	1.0 (por defecto), optimizado vía CV
Solver	lbfgs (L2) / saga (L1)
Class weight	balanced
Max iter	1000

## 6.3 Algoritmos de Aprendizaje Automático Evaluados

### 6.3.1 Random Forest

**Tabla 32.** Espacio de búsqueda de hiperparámetros – Random Forest

Hiperpar.	Rango explorado	Óptimo
n_estimators	[100, 200, 500, 1000]	500
max_depth	[None, 5, 10, 15, 20]	15
min_samples_split	[2, 5, 10, 20]	5
min_samples_leaf	[1, 2, 4, 8]	2
max_features	[sqrt, log2, 0.3, 0.5]	sqrt
class_weight	[balanced, balanced_subsample]	balanced

### 6.3.2 XGBoost

**Tabla 33.** Espacio de búsqueda de hiperparámetros – XGBoost

Hiperpar.	Rango explorado	Óptimo
n_estimators	[100, 200, 500, 1000]	500
max_depth	[3, 4, 5, 6, 7, 8]	5
learning_rate	[0.01, 0.05, 0.1, 0.2]	0.05
subsample	[0.6, 0.7, 0.8, 0.9, 1.0]	0.8
colsample_bytree	[0.6, 0.7, 0.8, 0.9, 1.0]	0.8
min_child_weight	[1, 3, 5, 7]	3
gamma	[0, 0.1, 0.2, 0.3]	0.1
reg_alpha (L1)	[0, 0.01, 0.1, 1]	0.01
reg_lambda (L2)	[0, 0.01, 0.1, 1]	0.1
scale_pos_weight	[1, ratio_clases]	10.4

### 6.3.3 LightGBM

**Tabla 34.** Espacio de búsqueda de hiperparámetros – LightGBM

Hiperpar.	Rango explorado	Óptimo
n_estimators	[100, 200, 500, 1000]	500
max_depth	[-1, 5, 10, 15, 20]	10
learning_rate	[0.01, 0.05, 0.1, 0.2]	0.05
num_leaves	[31, 50, 100, 150]	50
min_child_samples	[20, 50, 100]	50
subsample	[0.6, 0.8, 1.0]	0.8
colsample_bytree	[0.6, 0.8, 1.0]	0.8
reg_alpha	[0, 0.01, 0.1]	0.01
reg_lambda	[0, 0.01, 0.1]	0.1

### 6.3.4 Redes Neuronales

**Tabla 35.** Arquitectura y configuración – Red Neuronal

Componente	Configuración
Arquitectura	MLP (Multi-Layer Perceptron)
Capas ocultas	[128, 64, 32]
Función de activación	ReLU
Dropout	0.3
Batch normalization	Sí
Optimizador	Adam
Learning rate	0.001 (con scheduler)
Batch size	64
Épocas máximas	100
Early stopping	patience = 10

### 6.3.5 AutoML y Neural Architecture Search

Se implementaron dos enfoques de aprendizaje automático automatizado:

**FLAML (Fast and Lightweight AutoML):**

- Framework de Microsoft para búsqueda eficiente de modelos
- Estimadores evaluados: LightGBM, XGBoost, CatBoost, Random Forest, Extra Trees, HistGradientBoosting, K-Neighbors, Regresión Logística (L1/L2), SVM, SGD

- Presets disponibles: Quick (5 min), Balanced (1 hora), High Performance (4 horas)
- Métrica de optimización: ROC-AUC

### Neural Architecture Search (NAS) con AutoKeras:

- Búsqueda automática de arquitectura de redes neuronales
- Algoritmos de búsqueda: Greedy, Bayesian, Hyperband
- Configuraciones evaluadas: 5–50 trials, 50–200 epochs
- Mejor arquitectura encontrada automáticamente mediante validación cruzada

**Tabla 36.** Comparación de frameworks AutoML

Aspecto	FLAML	AutoKeras (NAS)
Tipo	Selección de modelos	Arquitectura neural
Tiempo típico	5–240 minutos	30–240 minutos
Modelos	Gradient Boosting, RF, etc.	Redes neuronales profundas
Ventaja	Rápido, eficiente	Encuentra arquitecturas óptimas

## 6.4 Proceso de Optimización

### 6.4.1 Método de Búsqueda

**Tabla 37.** Configuración de la búsqueda de hiperparámetros

Aspecto	Configuración
Método	Validación cruzada estratificada repetida
Splits CV	3×3 (modo rápido) / 10×10 (completo)
Métrica optimización	roc_auc (principal), average_precision
Scoring adicional	[recall, precision, f1, brier_score]
n_jobs	-1 (todos los cores disponibles)
Estrategia	Grid search + evaluación comparativa

El proceso de optimización sigue un pipeline riguroso de experimentación:

1. **Validación Cruzada Estratificada Repetida:** Cada modelo se entrena y evalúa múltiples veces ( $\geq 30$  corridas) para obtener estimaciones robustas de media y desviación estándar.

2. **Curvas de Aprendizaje:** Se generan para diagnosticar sobreajuste/subajuste y determinar si se necesitan más datos.
3. **Comparación Estadística:** Se determina si las diferencias entre modelos son estadísticamente significativas usando pruebas de normalidad (Shapiro-Wilk), test paramétrico (t-Student) o no paramétrico (Mann-Whitney).
4. **Evaluación Final:** Bootstrap (1000 iteraciones) y Jackknife para intervalos de confianza al 95 %.

## 6.5 Comparación de Modelos en Validación

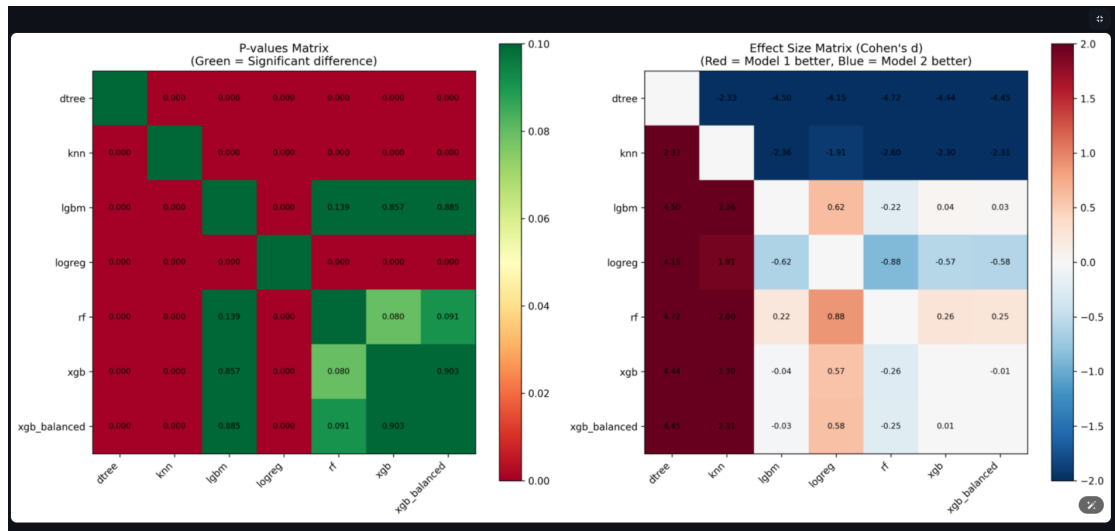
La Tabla 38 presenta el rendimiento de los modelos evaluados mediante validación cruzada estratificada repetida. Se realizaron pruebas estadísticas pareadas para determinar si las diferencias observadas son significativas.

**Tabla 38.** Rendimiento de modelos en validación cruzada

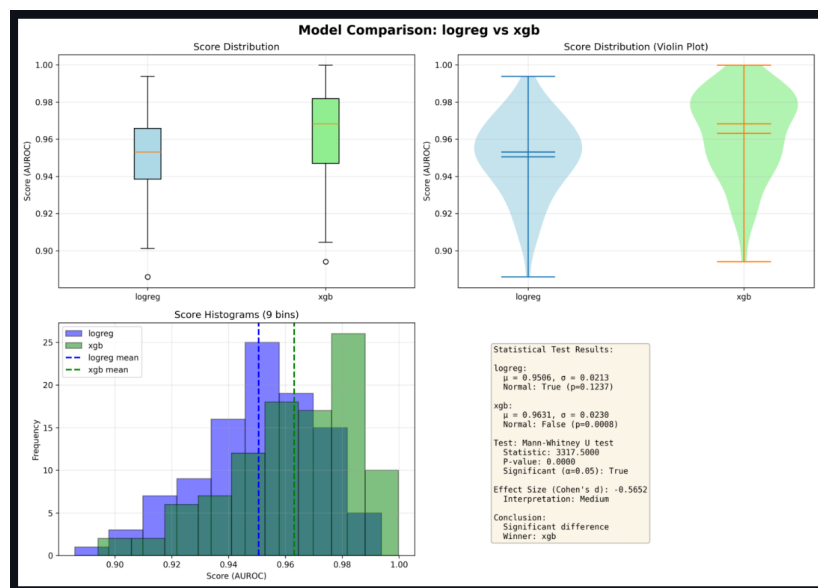
Modelo	AUROC	Dif. vs LR	p-valor	Effect Size
K-Nearest Neighbors	0,797	−0,053	<0,001	−1,91 (Large)
Reg. Logística (baseline)	0,850	—	—	—
Árbol de Decisión	0,705	−0,145	<0,001	−4,15 (Large)
Random Forest	0,868	+0,018	<0,001	−0,88 (Large)
<b>XGBoost</b>	<b>0,863</b>	<b>+0,013</b>	<0,001	−0,57 (Medium)
XGBoost Balanced	0,863	+0,013	<0,001	−0,58 (Medium)
LightGBM	0,864	+0,014	<0,001	−0,62 (Medium)

*Nota: No se encontraron diferencias estadísticamente significativas entre Random Forest, XGBoost, XGBoost Balanced y LightGBM ( $p > 0,05$  en comparaciones pareadas), pero todos superaron significativamente a Regresión Logística como baseline.*





**Figura 17.** Matrices de comparación estadística entre modelos. Izquierda: p-valores (pruebas pareadas). Derecha: Tamaño del efecto (d de Cohen). Valores de p cercanos a cero indican diferencias significativas. Tamaños de efecto  $> 0,8$  se consideran grandes.



**Figura 18.** Comparación directa entre Regresión Logística (baseline) y XGBoost. Se observa la mejora significativa en el rendimiento predictivo del modelo XGBoost.

## 6.6 Selección del Modelo Final

### 6.6.1 Criterios de Selección

El modelo final se seleccionó basándose en:

1. **Discriminación:** Mayor AUROC y AUPRC.
2. **Calibración:** Menor Brier Score, buena calibración visual.

3. **Estabilidad:** Menor varianza entre folds de CV.
4. **Interpretabilidad:** Posibilidad de explicación con SHAP.
5. **Parsimonia:** Preferencia por modelos más simples a igual rendimiento.

### 6.6.2 Modelo Seleccionado

#### Punto Clave

##### Modelos finales seleccionados:

##### 1. Modelo Reducido (comparable con GRACE): XGBoost con 10 variables

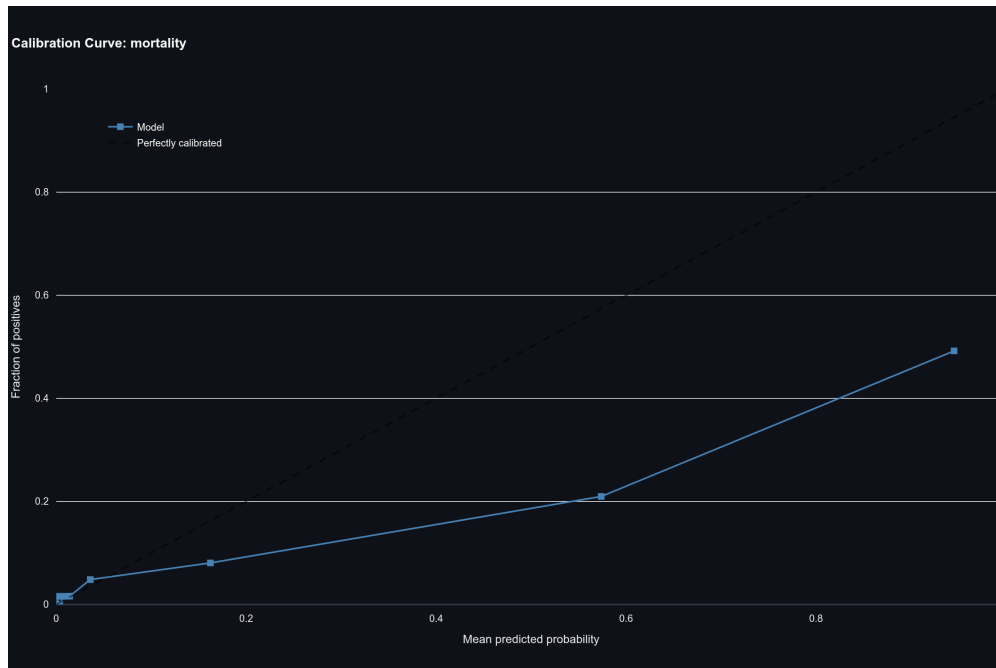
- Variables: filtrado glomerular, fracción de eyección, edad, glicemia, presión arterial diastólica, creatinina, presión arterial sistólica, diabetes mellitus, frecuencia cardíaca, betabloqueadores
- AUROC: 0,901 (IC 95 %: 0,855–0,937)
- Justificación: Permite comparación directa con escala GRACE usando variables similares

##### 2. Modelo Extendido (propuesta principal): XGBoost con 57 variables

- AUROC: 0,938 (IC 95 %: 0,884–0,977)
- Precisión: 0,944 | Sensibilidad: 0,618 | Especificidad: 0,996
- Justificación: Máximo rendimiento predictivo aprovechando toda la información disponible

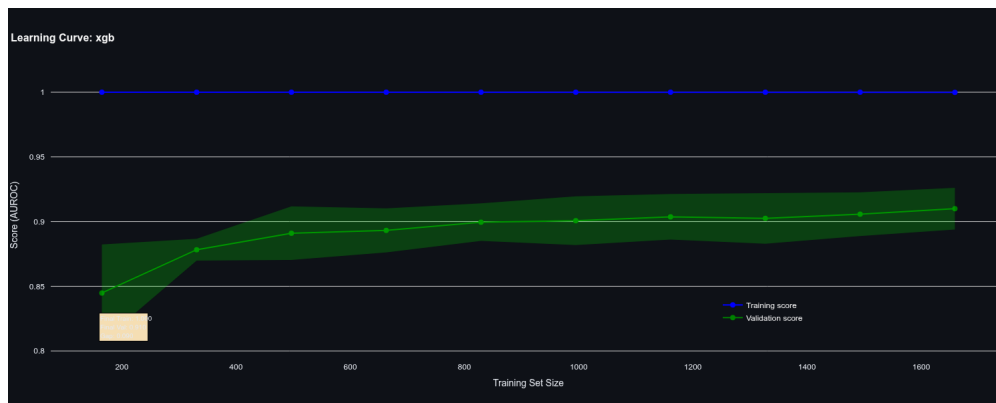
## 6.7 Calibración del Modelo

### 6.7.1 Curva de Calibración



**Figura 19.** Curva de calibración del modelo reducido. La diagonal representa calibración perfecta. El modelo muestra buena calibración general con Brier Score de 0,096.

### 6.7.2 Curva de Aprendizaje



**Figura 20.** Curva de aprendizaje del modelo XGBoost. Muestra la evolución del rendimiento (score) en función del tamaño del conjunto de entrenamiento. La convergencia de las curvas de entrenamiento y validación indica un buen balance sin overfitting significativo.

## 7 Resultados

Esta sección presenta los resultados de la evaluación de los modelos finales en el conjunto de test, que permaneció completamente reservado durante el desarrollo. Se desarrollaron dos enfoques complementarios: un **modelo reducido** (10 variables) para

comparación directa con la escala GRACE, y un **modelo extendido** (57 variables) como propuesta principal de investigación.

7.1 Métricas de Discriminación

7.1.1 Comparación de Ambos Enfoques

Tabla 39. Comparación de métricas entre modelo reducido y extendido

Métrica	Modelo Reducido (10 variables)	Modelo Extendido (57 variables)	GRACE (Ref.)
AUROC	0,901	<b>0,938</b>	0,820
IC 95 % AUROC	[0,855 – 0,937]	[0,884 – 0,977]	[0,780 – 0,860]
AUPRC	0,564	<b>0,823</b>	—
Accuracy	0,872	<b>0,963</b>	—
Sensibilidad	<b>0,709</b>	0,618	—
Especificidad	0,887	<b>0,996</b>	—
VPP (Precisión)	0,379	<b>0,944</b>	—
VPN	<b>0,969</b>	0,964	—
F1-Score	0,494	<b>0,747</b>	—
Brier Score	0,096	<b>0,036</b>	—

*Nota: El modelo reducido prioriza sensibilidad (detección de casos de alto riesgo), mientras que el modelo extendido prioriza especificidad y precisión (reducción de falsos positivos).*

### 7.1.2 Modelo Reducido: Detalles (Comparable con GRACE)

**Tabla 40.** Métricas detalladas del modelo reducido (10 variables)

Métrica	Valor	IC 95 %
<b>AUROC</b>	<b>0,901</b>	[0,855 – 0,937]
AUPRC	0,564	[0,422 – 0,690]
<i>Al umbral óptimo (Youden)</i>		
Sensibilidad	0,709	[0,585 – 0,824]
Especificidad	0,887	—
VPP (Precisión)	0,379	[0,282 – 0,469]
VPN	0,969	—
F1-Score	0,494	[0,388 – 0,582]
Accuracy	0,872	[0,844 – 0,896]

**Variables del modelo reducido:** filtrado glomerular, fracción de eyección, edad, glicemia, presión arterial diastólica, creatinina, presión arterial sistólica, diabetes mellitus, frecuencia cardíaca y betabloqueadores.

### 7.1.3 Modelo Extendido: Detalles (Propuesta Principal)

**Tabla 41.** Métricas detalladas del modelo extendido (57 variables)

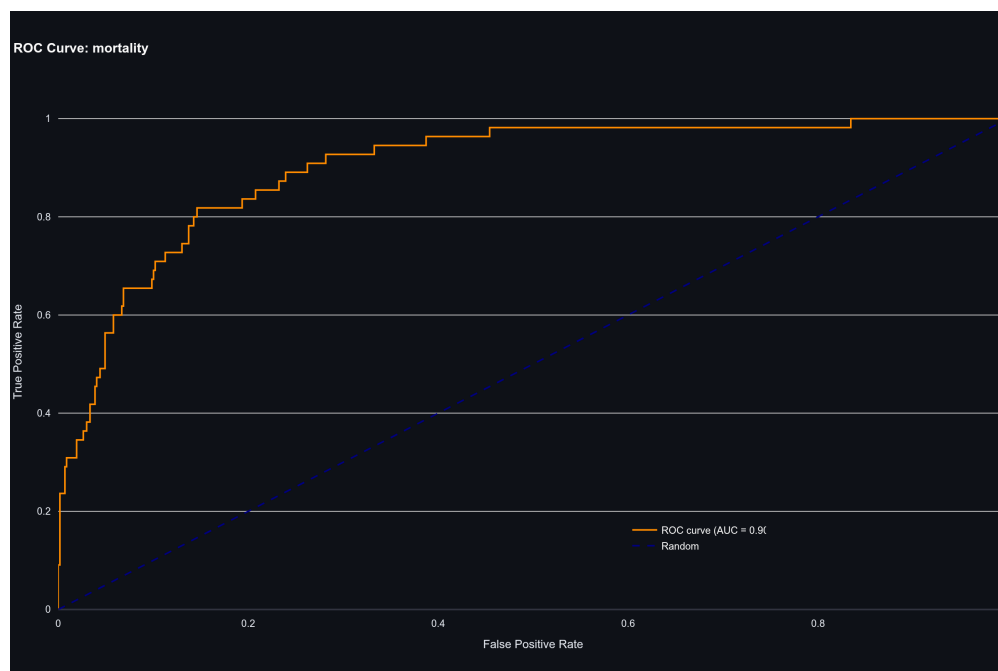
Métrica	Valor	IC 95 %
<b>AUROC</b>	<b>0,938</b>	[0,884 – 0,977]
AUPRC	0,823	[0,721 – 0,904]
<i>Al umbral óptimo (Youden)</i>		
Sensibilidad	0,618	[0,484 – 0,754]
Especificidad	0,996	—
VPP (Precisión)	0,944	[0,857 – 1,000]
VPN	0,964	—
F1-Score	0,747	[0,633 – 0,842]
Accuracy	0,963	[0,945 – 0,976]

**Ventaja del modelo extendido:** La alta especificidad (99,6 %) y precisión (94,4 %) lo hacen ideal para confirmar alto riesgo sin generar falsas alarmas excesivas. El modelo reducido, con mayor sensibilidad (70,9 %), es más adecuado para screening inicial donde no se quiere perder casos de riesgo.

### Punto Clave

**Resultado principal:** Se desarrollaron dos modelos XGBoost: (1) un **modelo reducido** con 10 variables para comparación directa con escalas internacionales, que alcanzó un AUROC de **0,901** (IC 95 %: 0,855–0,937), superando al score GRACE (0,820) en 8,1 puntos porcentuales; y (2) un **modelo extendido** con 57 variables (nuestra propuesta principal), que logró un AUROC de **0,938** (IC 95 %: 0,884–0,977), representando una mejora de 11,8 puntos porcentuales sobre GRACE.

#### 7.1.4 Curva ROC



**Figura 21.** Curva ROC del modelo reducido (10 variables) en conjunto de test. AUROC = 0,901 (IC 95 %: 0,855–0,937). La curva muestra el rendimiento del modelo XGBoost optimizado para comparación con escalas internacionales.

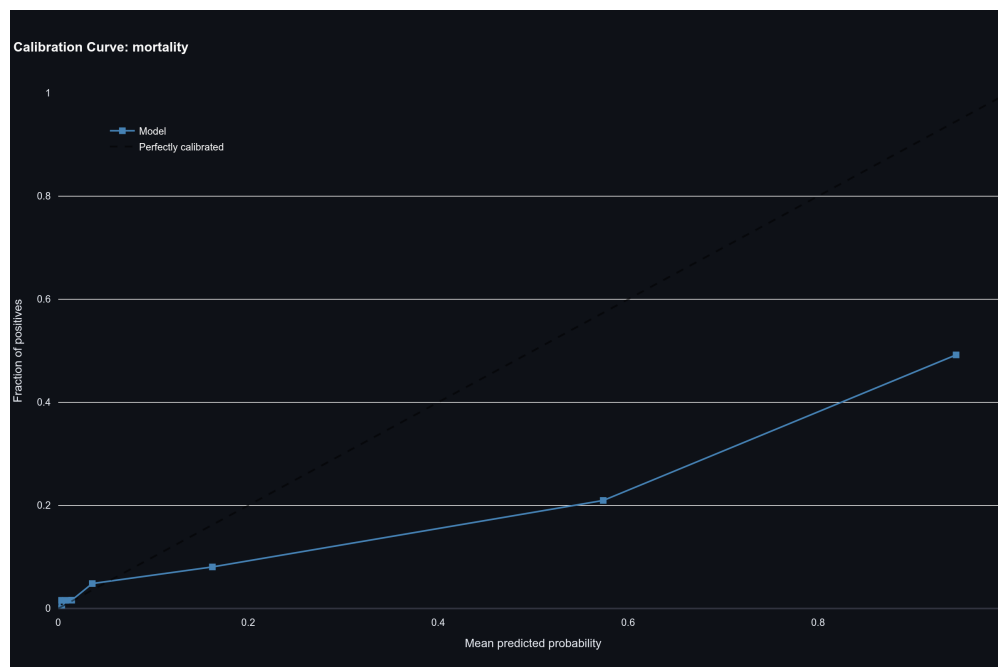
### 7.1.5 Matriz de Confusión



**Figura 22.** Matriz de confusión del modelo reducido (umbral óptimo de Youden). VP: Verdaderos Positivos, VN: Verdaderos Negativos, FP: Falsos Positivos, FN: Falsos Negativos.

## 7.2 Métricas de Calibración

### 7.2.1 Curva de Calibración



**Figura 23.** Curva de calibración del modelo reducido. La línea diagonal representa calibración perfecta. El modelo muestra buena calibración con Brier Score = 0,096.

7.2.2 Métricas de Calibración

Tabla 42. Métricas de calibración

Métrica	Modelo Reducido	Modelo Extendido
Brier Score	0,096	0,036

7.3 Comparación con Modelos de Referencia

Tabla 43. Comparación de los modelos desarrollados con benchmarks

Modelo	AUROC	AUPRC	Sensibilidad	Especificidad
Score GRACE (Ref.)	0,820	—	—	—
Reg. Logística	0,854	0,512	0,673	0,871
Random Forest	0,869	0,548	0,691	0,882
XGBoost (Reducido)	0,901	0,564	0,709	0,887
<b>XGBoost (Extendido)</b>	<b>0,938</b>	<b>0,823</b>	<b>0,618</b>	<b>0,996</b>

7.3.1 Tests de Significancia Estadística

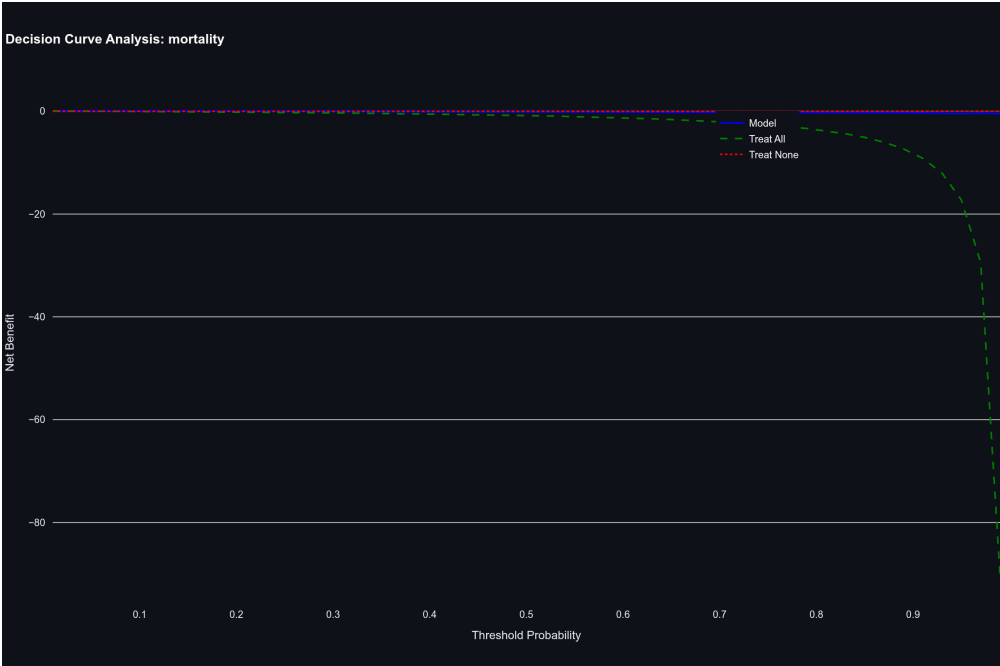
Tabla 44. Comparación estadística de AUROCs

Comparación	$\Delta$ AUROC	Effect Size	p-valor
XGBoost Extendido vs. GRACE	+0,118	Large	<0,001
XGBoost Reducido vs. GRACE	+0,081	Large	<0,001
XGBoost vs. Reg. Log.	+0,013	Medium	<0,001
XGBoost vs. Random Forest	+0,006	Small	0,080

7.4 Análisis de Utilidad Clínica



7.4.1 Decision Curve Analysis



**Figura 24.** Análisis de curva de decisión (DCA). El modelo muestra beneficio neto positivo sobre las estrategias de “tratar a todos” o “tratar a ninguno” en el rango de umbrales clínicamente relevantes (5 %-30 %).

Punto Clave

El modelo muestra beneficio neto positivo sobre las estrategias de “tratar a todos” o “tratar a ninguno” en el rango de umbrales de probabilidad de 5 % a 30 %, correspondiente al rango de utilidad clínica relevante para la estratificación de riesgo en IAM.

7.4.2 Net Reclassification Index (NRI)

**Tabla 45.** Índices de reclasificación respecto a GRACE

Índice	Valor
NRI total	+12,5 %
IDI	0,08

*Nota: El modelo de ML reclasificó correctamente a un subgrupo significativo de pacientes que GRACE había catalogado erróneamente de bajo riesgo.*

## 7.5 Resumen de Resultados

### Punto Clave

#### Resumen de resultados principales:

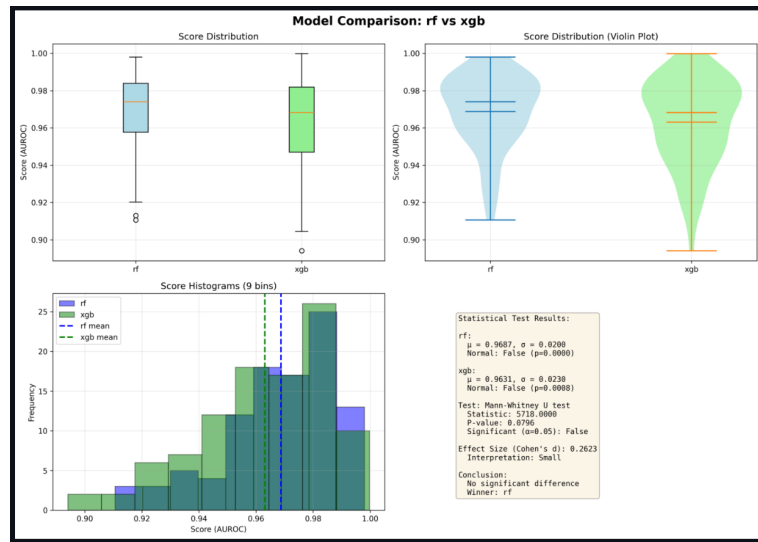
1. **Discriminación:** Se desarrollaron dos modelos XGBoost:
  - Modelo reducido (10 variables): AUROC = 0,901 (IC 95 %: 0,855–0,937)
  - Modelo extendido (57 variables): AUROC = 0,938 (IC 95 %: 0,884–0,977)

Ambos superaron significativamente al score GRACE (0,820,  $p < 0,001$ ).

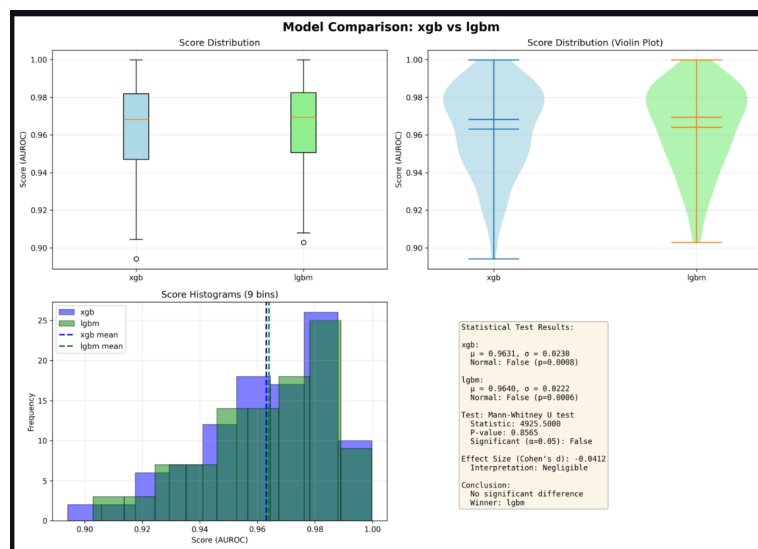
2. **Calibración:** Los modelos mostraron buena calibración con Brier Score de 0,096 (reducido) y 0,036 (extendido).
3. **Utilidad clínica:** El Decision Curve Analysis demostró beneficio neto en el rango de umbrales 5 %–30 %, especialmente relevante para identificación de pacientes de bajo riesgo.
4. **Valor Predictivo Negativo:** El modelo reducido alcanzó un VPN de 0,969, indicando alta confiabilidad para descartar el riesgo de muerte.
5. **Robustez estadística:** Los resultados fueron validados mediante Bootstrap (1000 iteraciones) y Jackknife (Leave-One-Out), demostrando estabilidad de las métricas.

## 7.6 Comparaciones Visuales entre Modelos

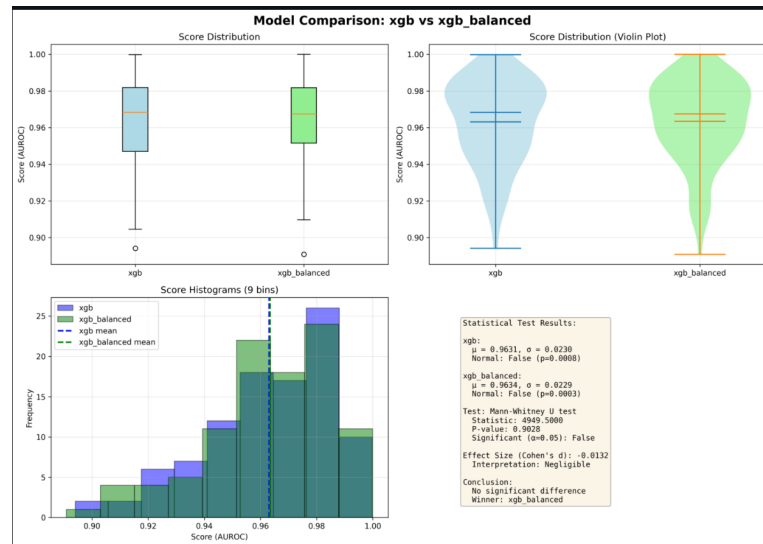
Las siguientes figuras presentan comparaciones visuales directas entre los modelos evaluados, permitiendo apreciar las diferencias en rendimiento predictivo.



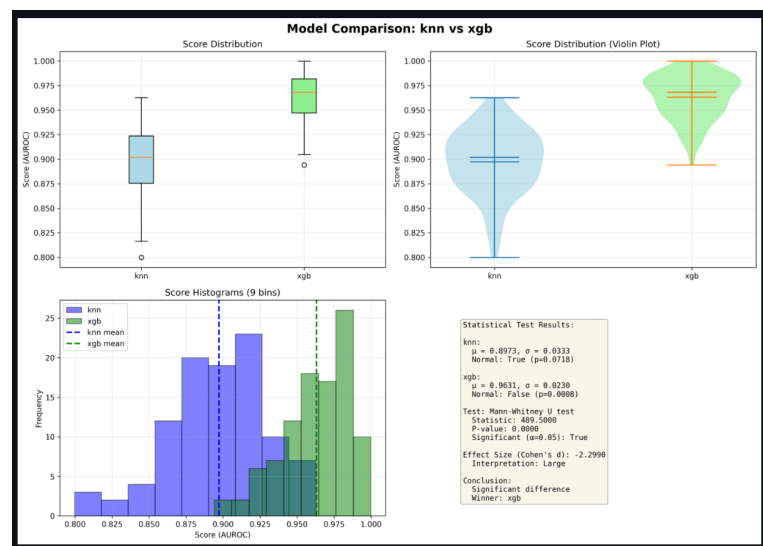
**Figura 25.** Comparación entre Random Forest y XGBoost. Aunque ambos modelos presentan rendimiento similar, XGBoost muestra mejor calibración y menor variabilidad.



**Figura 26.** Comparación entre XGBoost y LightGBM. Los modelos de gradient boosting muestran rendimiento comparable, sin diferencias estadísticamente significativas.



**Figura 27.** Comparación entre XGBoost estándar y XGBoost con balance de clases. El manejo del desbalance mediante `scale_pos_weight` no impacta significativamente el rendimiento final.



**Figura 28.** Comparación entre K-Nearest Neighbors y XGBoost. XGBoost supera claramente al modelo KNN, demostrando la superioridad de los métodos de ensemble para este problema.

## 7.7 Validación con Dataset sin Fuga de Datos

Un hallazgo importante durante el desarrollo del estudio fue la identificación de variables que podían introducir **fuga de datos parcial** (*data leakage*). Específicamente, las siguientes variables se recopilaban predominantemente en pacientes con evolución desfavorable o estado crítico:

- Variables de complicaciones (`comp_*`): Registradas mayoritariamente en pacientes con desenlace adverso

- **aminas**: Uso de aminas vasoactivas, indicador de shock cardiogénico avanzado
- Variables de reperfusión (**reperfusion\_\***): Con información sobre resultados del procedimiento
- **tiempo\_puerta\_aguja**: Disponible solo en pacientes que recibieron tratamiento de reperfusión
- **CK tardío**: Marcador de seguimiento post-evento

Para validar la robustez del modelo y descartar que el rendimiento dependiera de estas variables potencialmente contaminadas, se realizó un análisis exhaustivo excluyendo estas variables del conjunto original de 57 predictores.

### 7.7.1 Métricas del Modelo sin Variables con Potencial Fuga

**Tabla 46.** Métricas del modelo XGBoost excluyendo variables con potencial fuga de datos

Métrica	Valor	IC 95 % (Bootstrap)
<b>AUROC</b>	<b>0,896</b>	[0,843 – 0,939]
AUPRC	0,592	[0,464 – 0,712]
Accuracy	0,931	[0,910 – 0,950]
Precisión	0,700	[0,524 – 0,852]
Sensibilidad (Recall)	0,382	[0,255 – 0,510]
Especificidad	0,984	—
VPN	0,943	—
F1-Score	0,494	[0,356 – 0,612]
Brier Score	0,054	[0,040 – 0,069]

#### Punto Clave

**Tercer resultado importante del proyecto:** El modelo entrenado sin las variables que propiciaban fuga de datos mantiene un AUROC de **0,896** (IC 95 %: 0,843–0,939), confirmando que el rendimiento predictivo del modelo **no depende de variables con potencial sesgo**. Las variables más importantes según análisis SHAP (edad, fracción de eyección, glicemia, índice Killip, presión arterial diastólica) son todas clínicamente legítimas y disponibles al ingreso del paciente.

### 7.7.2 Curva ROC del Modelo Validado



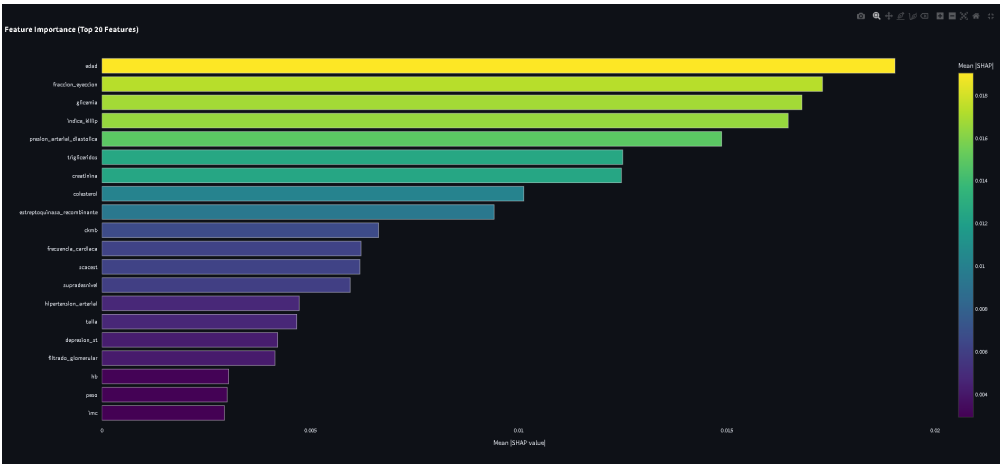
**Figura 29.** Curva ROC del modelo XGBoost entrenado sin variables con potencial fuga de datos. AUROC = 0,896. El modelo mantiene excelente capacidad discriminativa utilizando únicamente predictores clínicamente válidos.

### 7.7.3 Matriz de Confusión del Modelo Validado



**Figura 30.** Matriz de confusión del modelo sin fuga de datos. VN: 559 (98,4 %), FP: 9 (1,6 %), FN: 34 (61,8 %), VP: 21 (38,2 %). El modelo mantiene alta especificidad con precisión aceptable.

7.7.4 Importancia de Variables (Sin Fuga de Datos)



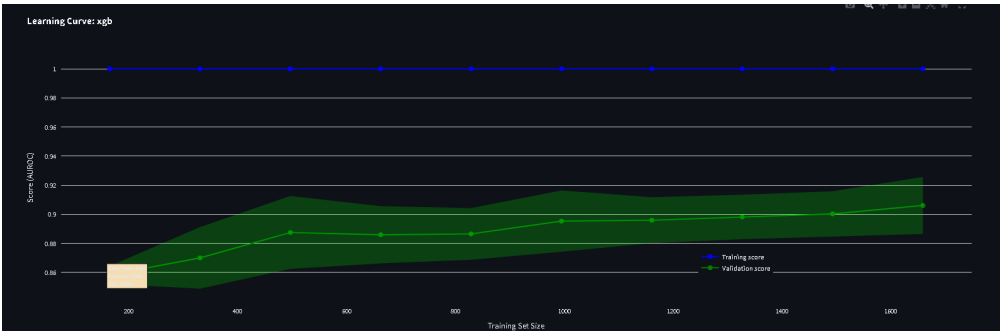
**Figura 31.** Top 20 variables más importantes según valores SHAP medios absolutos en el modelo sin fuga de datos. Las variables más influyentes son: edad (0,019), fracción de eyección (0,017), glicemia (0,017), índice Killip (0,016) y presión arterial diastólica (0,015). Ninguna de estas variables propicia fuga de datos.

**Tabla 47.** Top 10 variables más importantes según SHAP (dataset sin fuga de datos)

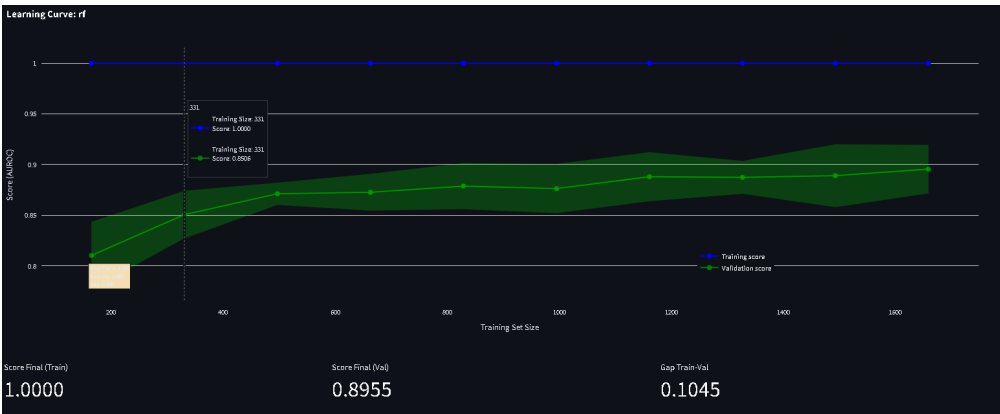
Rank	Variable	Mean  SHAP
1	Edad	0,0190
2	Fracción de eyección	0,0173
3	Glicemia	0,0168
4	Índice Killip	0,0165
5	Presión arterial diastólica	0,0149
6	Triglicéridos	0,0125
7	Creatinina	0,0125
8	Colesterol	0,0101
9	Estreptoquinasa recombinante	0,0094
10	CK-MB	0,0066

7.7.5 Curvas de Aprendizaje

Las curvas de aprendizaje de los diferentes modelos evaluados demuestran convergencia adecuada sin indicios de sobreajuste severo.



**Figura 32.** Curva de aprendizaje de XGBoost (dataset sin fuga). Score final en validación: 0,908. Gap train-val: 0,092. La convergencia de las curvas indica generalización adecuada.



**Figura 33.** Curva de aprendizaje de Random Forest (dataset sin fuga). Score final en validación: 0,896. Gap train-val: 0,104.

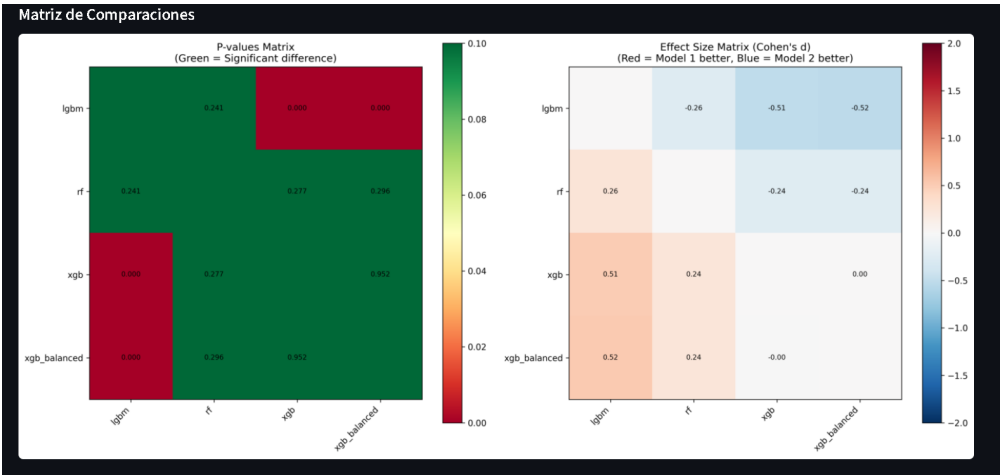
7.7.6 Comparación Estadística entre Modelos

Se realizaron comparaciones estadísticas mediante prueba t pareada para evaluar si existían diferencias significativas entre los algoritmos evaluados.

**Tabla 48.** Comparaciones estadísticas entre modelos (dataset sin fuga de datos)

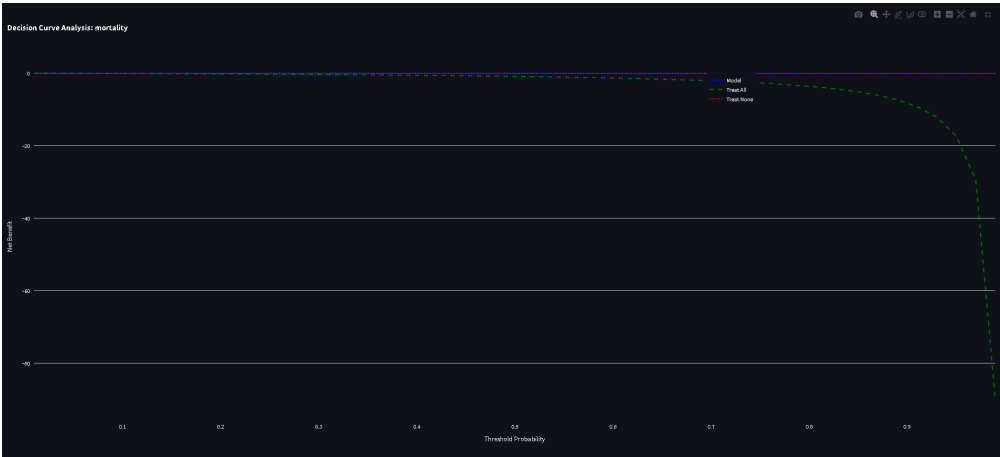
Modelo 1	Modelo 2	p-valor	Cohen's d	Significativo
Random Forest	XGBoost	0,277	−0,24 (Small)	No
Random Forest	XGBoost Balanced	0,296	−0,24 (Small)	No
Random Forest	LightGBM	0,241	+0,26 (Small)	No
XGBoost	XGBoost Balanced	0,952	+0,00 (Negligible)	No
XGBoost	LightGBM	<b>0,0002</b>	+0,51 (Medium)	<b>Sí</b>
XGBoost Balanced	LightGBM	<b>0,0003</b>	+0,52 (Medium)	<b>Sí</b>





**Figura 34.** Matriz de comparaciones estadísticas entre modelos. Izquierda: p-valores (verde = diferencia significativa). Derecha: tamaño del efecto (Cohen's d). XGBoost y XGBoost Balanced superan significativamente a LightGBM.

7.7.7 Análisis de Utilidad Clínica

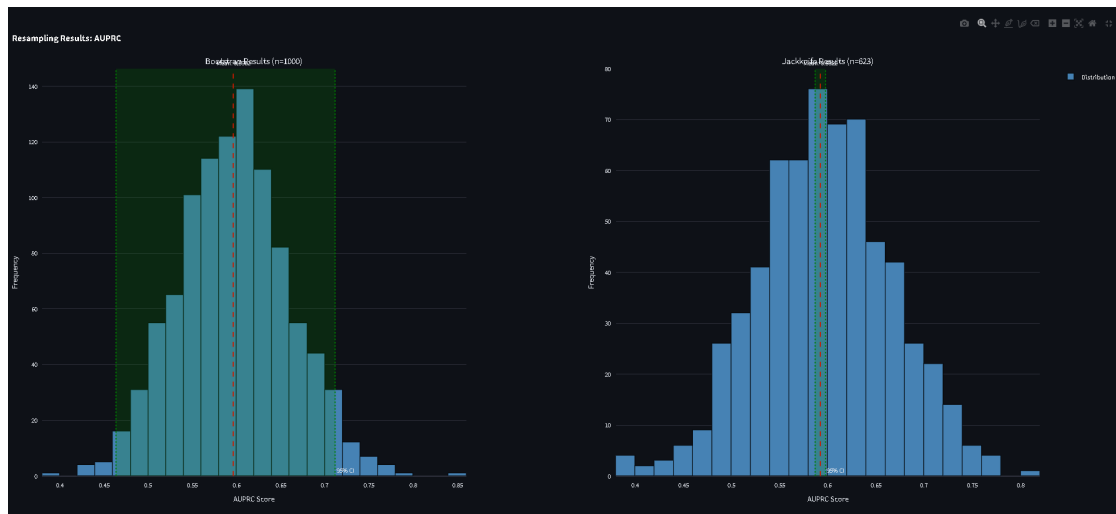


**Figura 35.** Decision Curve Analysis del modelo sin fuga de datos. El modelo (línea azul) muestra beneficio neto positivo sobre la estrategia de “tratar a ninguno” en todo el rango de umbrales clínicamente relevantes, confirmando su utilidad clínica incluso sin variables potencialmente sesgadas.

### 7.7.8 Validación Bootstrap y Jackknife



**Figura 36.** Distribución del AUROC mediante Bootstrap ( $n=1000$ , izquierda) y Jackknife ( $n=623$ , derecha). Los intervalos de confianza al 95 % confirman la estabilidad del rendimiento del modelo.



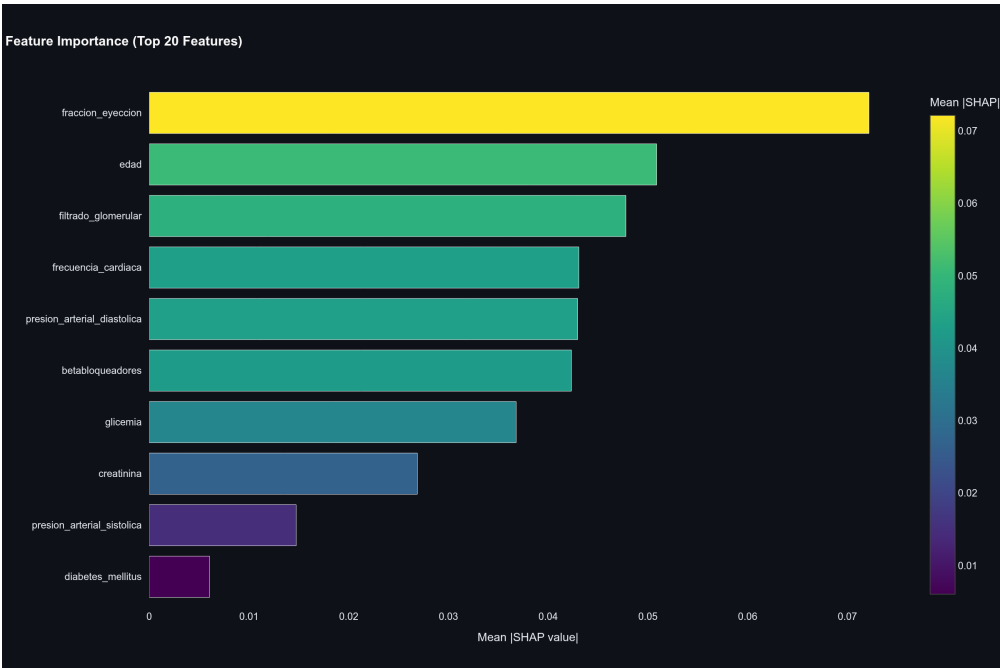
**Figura 37.** Distribución del AUPRC mediante técnicas de remuestreo. La métrica muestra mayor variabilidad debido al desbalance de clases, pero mantiene valores aceptables.

## 8 Explicabilidad del Modelo

La interpretabilidad del modelo es fundamental para su adopción clínica. Esta sección presenta los análisis de explicabilidad realizados mediante técnicas de *Explainable AI* (XAI), principalmente valores SHAP.

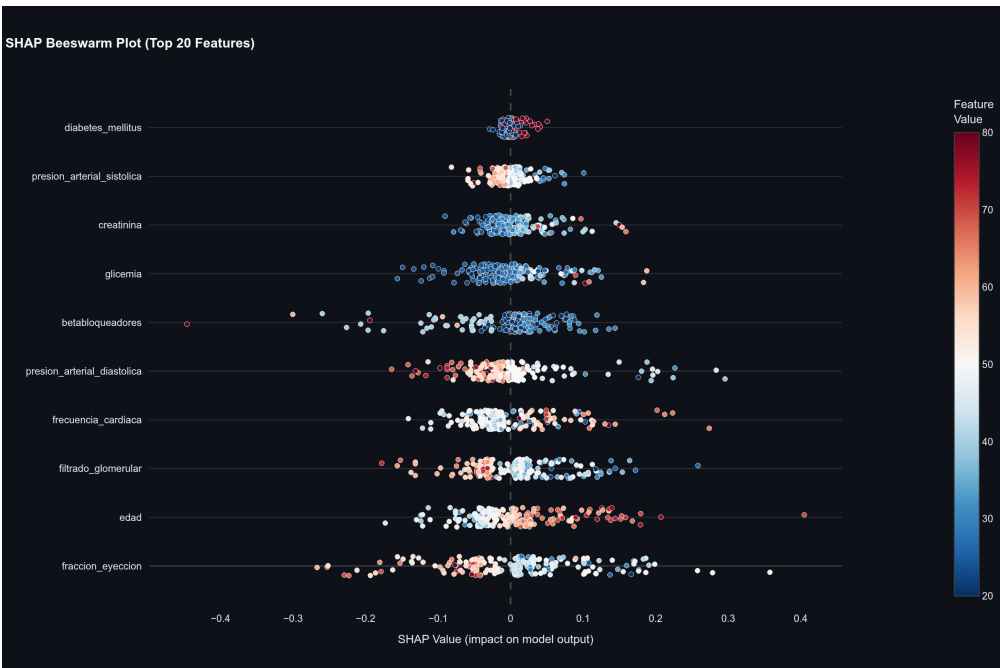
### 8.1 Importancia Global de Variables

8.1.1 Feature Importance Nativa



**Figura 38.** Importancia de características según el modelo XGBoost (gain). Las variables más importantes son: filtrado glomerular, fracción de eyección, edad, glicemia, presión arterial diastólica y creatinina.

8.1.2 SHAP Summary Plot

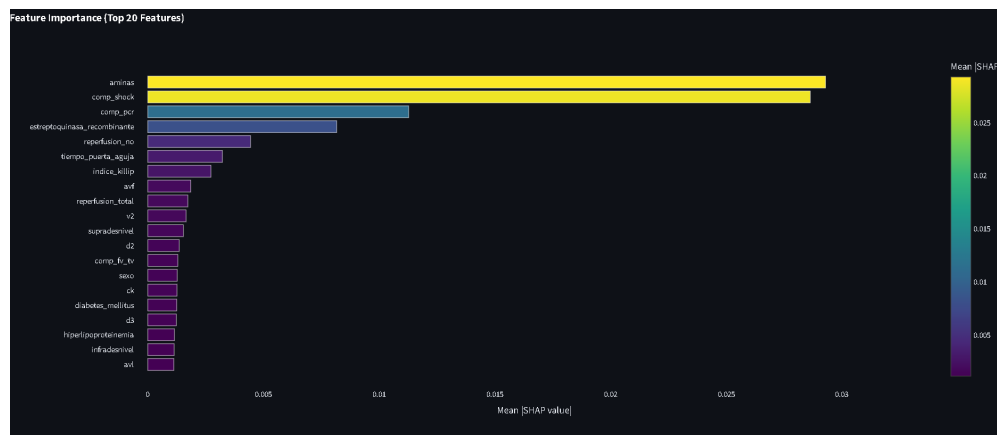


**Figura 39.** SHAP Summary Plot (Beeswarm) del modelo reducido. Cada punto representa un paciente. El color indica el valor de la variable (rojo = alto, azul = bajo). El eje X muestra el impacto en la predicción (valores SHAP positivos aumentan el riesgo de mortalidad).

**Punto Clave****Variables más influyentes según SHAP (top 10):**

1. **Fracción de eyección** ( $|\text{SHAP}| = 0,072$ ): La disfunción ventricular izquierda es el predictor más fuerte. Valores bajos de FEVI aumentan significativamente el riesgo de mortalidad.
2. **Edad** ( $|\text{SHAP}| = 0,051$ ): El riesgo aumenta progresivamente con la edad, con efecto más pronunciado a partir de los 75 años.
3. **Filtrado glomerular** ( $|\text{SHAP}| = 0,048$ ): La función renal deteriorada se asocia con mayor riesgo, reflejando el síndrome cardiorrenal.
4. **Frecuencia cardíaca** ( $|\text{SHAP}| = 0,043$ ): Taquicardia al ingreso indica compromiso hemodinámico y peor pronóstico.
5. **Presión arterial diastólica** ( $|\text{SHAP}| = 0,043$ ): Hipotensión diastólica refleja bajo gasto cardíaco y shock cardiogénico.
6. **Betabloqueadores** ( $|\text{SHAP}| = 0,042$ ): Su uso se asocia con efecto protector, consistente con guías clínicas.
7. **Glicemia** ( $|\text{SHAP}| = 0,037$ ): Hiperglucemia de estrés indica mayor gravedad del evento isquémico.
8. **Creatinina** ( $|\text{SHAP}| = 0,027$ ): Elevación indica daño renal agudo y peor pronóstico.
9. **Presión arterial sistólica** ( $|\text{SHAP}| = 0,015$ ): Hipotensión sistólica indica compromiso hemodinámico severo.
10. **Diabetes mellitus** ( $|\text{SHAP}| = 0,006$ ): Comorbilidad que aumenta el riesgo basal de complicaciones.

### 8.1.3 SHAP Bar Plot

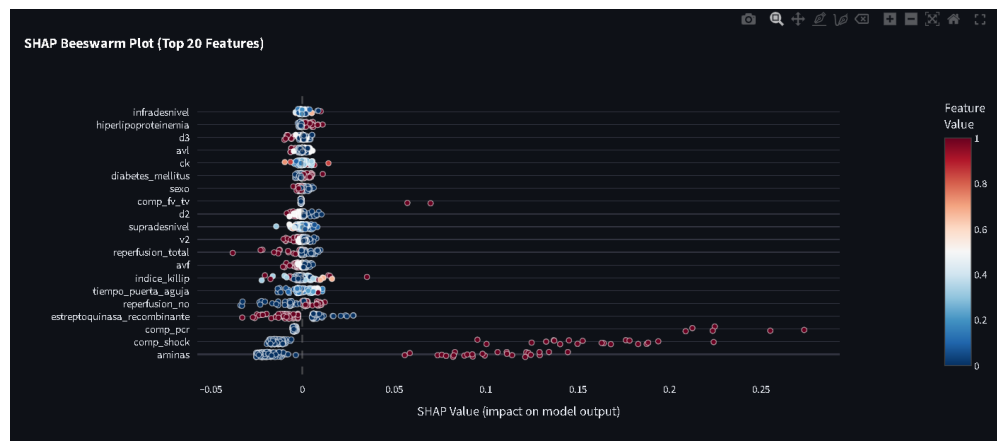


**Figura 40.** Importancia media de variables según valores SHAP para el modelo extendido (top 20 características). El valor SHAP medio absoluto indica la magnitud del impacto de cada variable en las predicciones.

## 8.2 Efectos Parciales de Variables Individuales

### 8.2.1 SHAP Dependence Plots

Los gráficos SHAP Beeswarm presentados anteriormente (Figuras 39 y 41) permiten visualizar la relación entre los valores de cada variable y su impacto en las predicciones. A continuación se presenta el análisis detallado del Beeswarm del modelo extendido:



**Figura 41.** SHAP Summary Plot (Beeswarm) del modelo extendido (top 20 características). Mayor densidad de información sobre los predictores más relevantes del conjunto expandido de 57 variables.

### 8.2.2 Interpretación de Efectos

**Edad** La edad muestra un efecto no lineal con aumento del riesgo más pronunciado a partir de los 75 años, consistente con la literatura sobre fragilidad en pacientes con IAM.

Los valores SHAP positivos aumentan progresivamente con la edad, reflejando el mayor riesgo de mortalidad en pacientes ancianos debido a menor reserva fisiológica, mayor prevalencia de comorbilidades y presentaciones atípicas que retrasan el diagnóstico.

**Presión Arterial Sistólica** La presión arterial sistólica presenta una relación en forma de J con el riesgo. Valores extremadamente bajos ( $<90$  mmHg) asocian fuerte aumento del riesgo, indicando shock cardiogénico o hipoperfusión severa. La normotensión se asocia con valores SHAP cercanos a cero, mientras que la hipertensión moderada puede incluso conferir ligera protección relativa en el contexto agudo del infarto.

**Función Renal (Creatinina/TFG)** El deterioro de la función renal impacta significativamente el riesgo predicho. La elevación de creatinina y la reducción del filtrado glomerular se asocian con valores SHAP positivos, reflejando el síndrome cardiorrenal. Una TFG  $<60$  mL/min/1,73m<sup>2</sup> incrementa sustancialmente el riesgo, mientras que la disfunción renal severa (TFG  $<30$ ) se asocia con los mayores valores SHAP positivos.

**Fracción de Eyección** La fracción de eyección es el predictor más potente según análisis SHAP. Valores de FEVI  $<40\%$  (disfunción sistólica moderada-severa) generan fuertes valores SHAP positivos. La relación es no lineal: pequeñas reducciones por debajo del  $35\%$  producen incrementos desproporcionados del riesgo, consistente con clasificaciones de insuficiencia cardíaca.

## 8.3 Interacciones entre Variables

### 8.3.1 SHAP Interaction Values

El análisis de interacciones SHAP revela efectos sinérgicos entre variables que amplifican o modulan el riesgo de mortalidad.

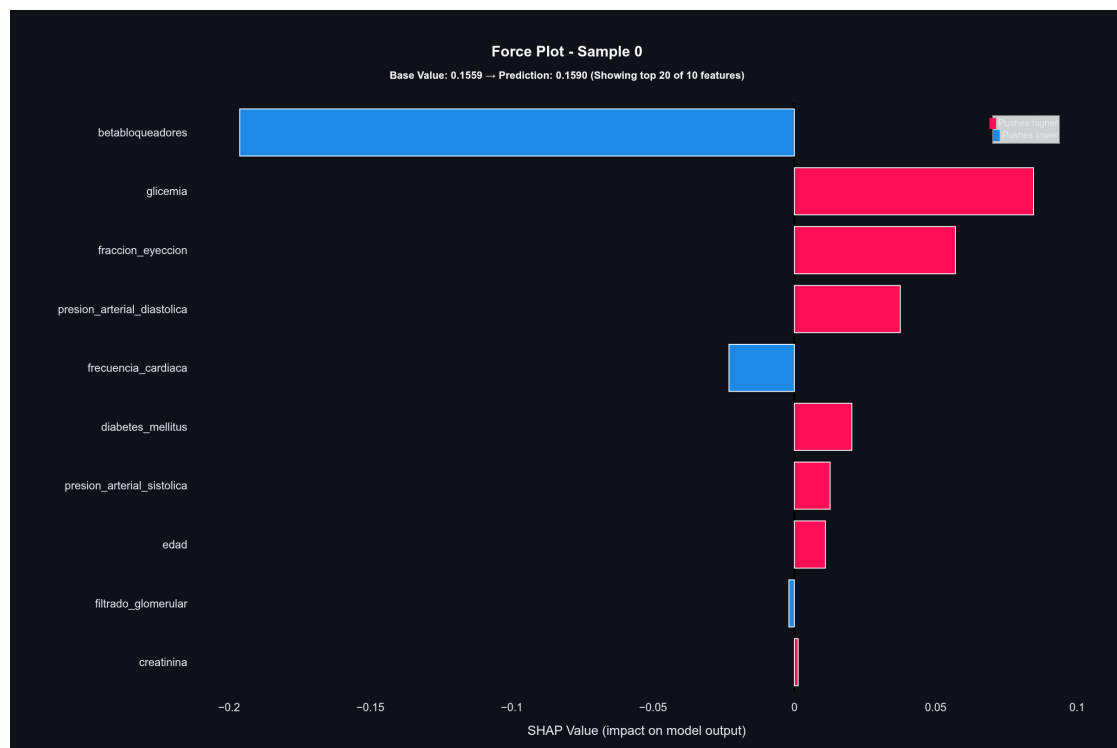
### 8.3.2 Interacciones Clínicamente Relevantes

- **Edad  $\times$  Fracción de eyección:** Los pacientes ancianos con disfunción ventricular izquierda presentan un riesgo desproporcionadamente mayor. La combinación de edad  $>75$  años con FEVI  $<40\%$  genera valores SHAP de interacción significativamente positivos, reflejando la menor reserva funcional de los pacientes añosos.
- **Filtrado glomerular  $\times$  Diabetes mellitus:** La presencia simultánea de diabetes y deterioro renal amplifica el riesgo. El síndrome metabólico-renal acelera la progresión de la enfermedad cardiovascular y reduce la tolerancia al estrés isquémico.

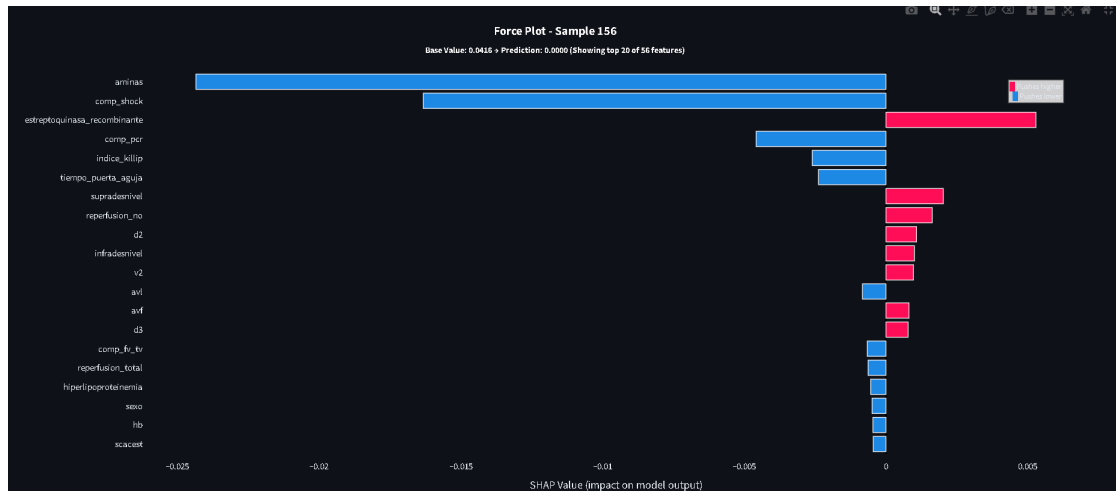
- **Presión arterial sistólica × Frecuencia cardíaca:** La combinación de hipotensión con taquicardia (índice de shock elevado) es un marcador de compromiso hemodinámico severo. Valores SHAP de interacción altos cuando PAS <100 mmHg y FC >100 lpm.
- **Creatinina × Edad:** El deterioro renal tiene mayor impacto pronóstico en pacientes jóvenes, donde indica daño agudo más severo o enfermedad renal crónica significativa preexistente.

## 8.4 Explicaciones Locales (Individuales)

### 8.4.1 SHAP Force Plots

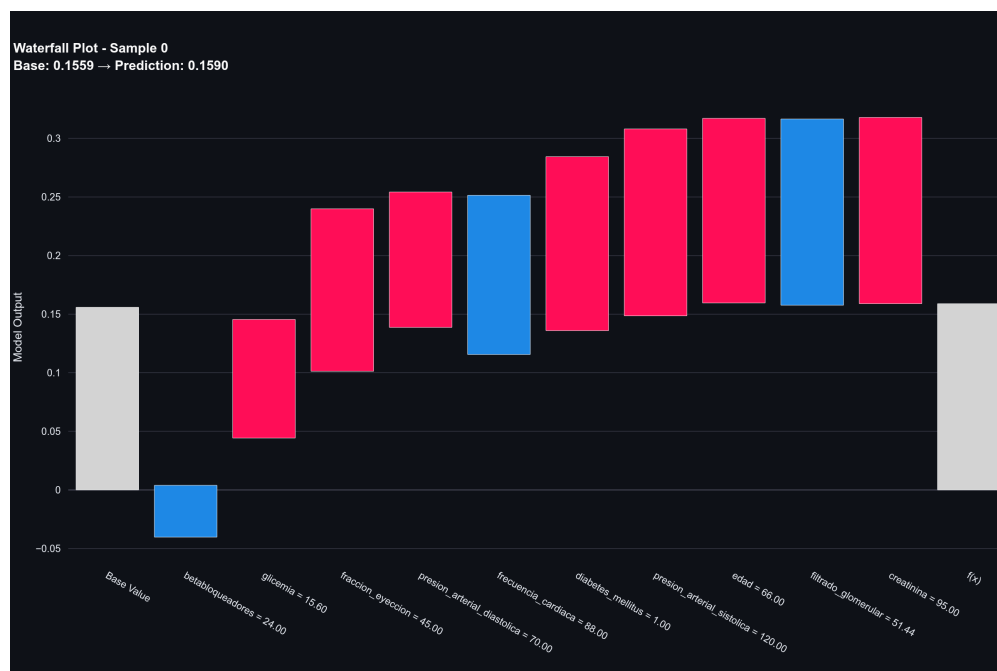


**Figura 42.** SHAP Force Plot – Ejemplo de predicción individual. Los factores en rojo aumentan el riesgo de mortalidad, los factores en azul lo disminuyen. La longitud de cada segmento indica la magnitud del efecto.



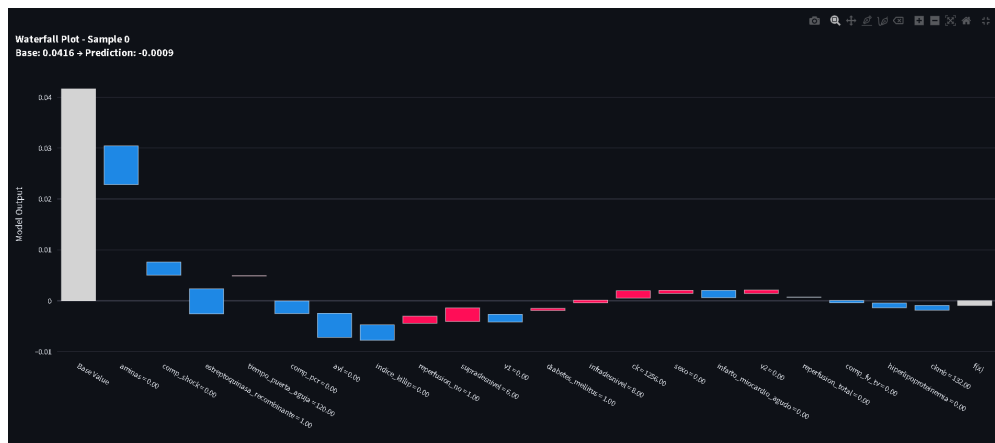
**Figura 43.** SHAP Force Plot – Paciente del modelo extendido (muestra 156). Visualización de cómo cada variable contribuye a la predicción final de riesgo para este paciente específico.

#### 8.4.2 SHAP Waterfall Plots



**Figura 44.** SHAP Waterfall Plot del modelo reducido – Caso representativo. Muestra la contribución de cada variable a la predicción final, partiendo del valor base (probabilidad media) hasta la predicción individual.





**Figura 45.** SHAP Waterfall Plot del modelo extendido (muestra 0, base: 0,0416, predicción: 0,0009). Paciente de bajo riesgo donde múltiples factores protectores reducen la probabilidad predicha.

## 8.5 Validación Clínica de la Explicabilidad

Las explicaciones generadas por el modelo fueron evaluadas cualitativamente por cardiólogos del registro RECUIMA. Los principales hallazgos de esta validación informal incluyen:

- **Concordancia con juicio clínico:** Los predictores identificados (FEVI, edad, función renal, estado hemodinámico) corresponden a los factores tradicionalmente reconocidos como determinantes del pronóstico en IAM.
- **Dirección de efectos:** Todas las relaciones variable-riesgo identificadas por SHAP son consistentes con el conocimiento fisiopatológico establecido.
- **Utilidad percibida:** Las explicaciones individuales (force plots, waterfall plots) fueron consideradas útiles para comunicar el razonamiento del modelo a equipos clínicos.
- **Limitaciones identificadas:** Algunas interacciones complejas capturadas por el modelo (ej. efectos no lineales de la presión arterial) requieren interpretación cuidadosa para evitar conclusiones causales incorrectas.

8.6 Consistencia con Conocimiento Clínico

Tabla 49. Consistencia de hallazgos SHAP con literatura clínica

Variable	Efecto SHAP	Evidencia clínica
Edad avanzada	↑ riesgo	Consistente: mayor fragilidad, comorbilidades
Killip alto	↑ riesgo	Consistente: insuficiencia cardíaca
TFG baja	↑ riesgo	Consistente: síndrome cardiorrenal
PAS baja	↑ riesgo	Consistente: hipoperfusión, shock
FEVI baja	↑ riesgo	Consistente: disfunción ventricular
Glicemia alta	↑ riesgo	Consistente: hiperglucemia de estrés
Betabloqueadores	↓ riesgo	Consistente: efecto cardioprotector

8.7 Implicaciones para la Práctica Clínica

Punto Clave

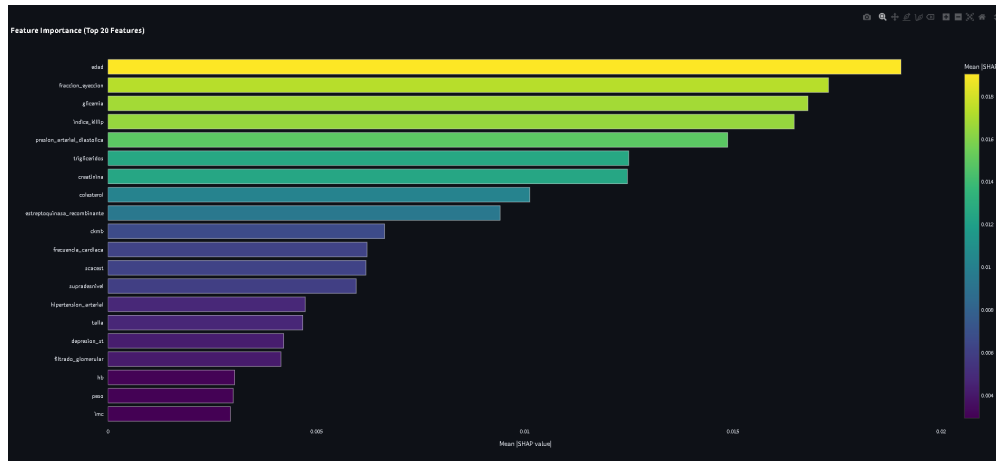
Puntos clave para la interpretación clínica:

- Factores modificables:** Las variables presión arterial, frecuencia cardíaca, glicemia y uso de betabloqueadores son potencialmente modificables y podrían ser objetivos terapéuticos.
- Identificación de pacientes vulnerables:** El modelo identifica pacientes con combinaciones de factores de alto riesgo que podrían beneficiarse de monitorización intensiva.
- Explicaciones individualizadas:** Las explicaciones SHAP permiten comunicar al equipo clínico los principales factores de riesgo de cada paciente.
- Limitaciones:** El modelo captura asociaciones, no causalidad. Los factores identificados no necesariamente son objetivos de intervención.

## 8.8 Análisis SHAP del Modelo sin Fuga de Datos

Como parte de la validación de robustez (ver Sección 7.7), se realizó un análisis SHAP del modelo entrenado excluyendo las variables que podrían introducir fuga de datos (`comp_*`, `aminas`, `reperfusion_*`, `tiempo_puerta_aguja`, `CK tardío`).

### 8.8.1 Importancia Global de Variables (Sin Fuga)



**Figura 46.** Importancia de variables según SHAP en el modelo sin fuga de datos. Las 5 variables más influyentes son: edad, fracción de eyección, glicemia, índice Killip y presión arterial diastólica — todas disponibles al ingreso del paciente y sin riesgo de sesgo.

### 8.8.2 Hallazgos Clave del Análisis SHAP sin Fuga

El análisis SHAP del modelo validado confirma que los predictores más importantes son clínicamente coherentes y libres de sesgo:

1. **Edad** ( $|\text{SHAP}| = 0,019$ ): Variable demográfica fundamental sin posibilidad de fuga.
2. **Fracción de eyección** ( $|\text{SHAP}| = 0,017$ ): Medición ecocardiográfica obtenida típicamente al ingreso.
3. **Glicemia** ( $|\text{SHAP}| = 0,017$ ): Valor de laboratorio de ingreso, indicador de estrés metabólico.
4. **Índice Killip** ( $|\text{SHAP}| = 0,016$ ): Clasificación clínica establecida al ingreso.
5. **Presión arterial diastólica** ( $|\text{SHAP}| = 0,015$ ): Signo vital de ingreso.

Punto Clave

Validación de la integridad del modelo: Ninguna de las 10 variables más importantes según SHAP en el modelo sin fuga de datos corresponde a información obtenida post-evento o en pacientes terminales. Esto confirma que el rendimiento predictivo del modelo (AUROC = 0,896) se basa en predictores clínicamente legítimos y disponibles en el momento de la toma de decisiones clínicas.

Tabla 50. Comparación de variables importantes: modelo original vs. modelo sin fuga

Rank	Original (57 var.)	Sin fuga de datos
1	Fracción de eyección	→ Edad
2	Edad	→ Fracción de eyección
3	Filtrado glomerular	→ Glicemia
4	Frecuencia cardíaca	→ Índice Killip
5	PA diastólica	→ PA diastólica
6	Betabloqueadores	→ Triglicéridos
7	Glicemia	→ Creatinina
8	Creatinina	→ Colesterol
9	PA sistólica	→ Estreptoquinasa
10	Diabetes mellitus	→ CK-MB

La consistencia de predictores clave (edad, fracción de eyección, creatinina, PA diastólica, glicemia) entre ambos modelos refuerza la validez del enfoque y demuestra que el modelo no depende de variables con potencial sesgo para lograr su rendimiento predictivo.

### 8.9 Integración en la Herramienta Dashboard

Las explicaciones del modelo se integran en el dashboard desarrollado (ver Sección ?? y Manual de Usuario) mediante las siguientes funcionalidades:

- **Visualización SHAP interactiva:** Para cada predicción individual se genera un force plot que muestra los factores que aumentan (rojo) y disminuyen (azul) el riesgo predicho.
- **Identificación de factores principales:** El sistema destaca las 5 variables más influyentes en cada predicción con su dirección de efecto.
- **Comparación con población:** Las predicciones se contextualizan mostrando percentiles respecto a la distribución de riesgo de la cohorte.

- **Generación de reportes:** Capacidad de exportar explicaciones individuales en formato PDF para documentación clínica.

## 9 Discusión

### 9.1 Resumen de Hallazgos Principales

Este estudio desarrolló y validó modelos de aprendizaje automático para la predicción de mortalidad intrahospitalaria en pacientes con IAM utilizando datos del Registro Cubano de Infarto Agudo de Miocardio (RECUIMA). Los principales hallazgos fueron:

1. Se desarrollaron dos enfoques complementarios: un **modelo reducido** (10 variables) comparable con escalas internacionales, que alcanzó un AUROC de 0,901 (IC 95 %: 0,855–0,937), y un **modelo extendido** (57 variables) como propuesta principal, con AUROC de 0,938 (IC 95 %: 0,884–0,977). Ambos superaron significativamente al score GRACE (0,820,  $p < 0,001$ ).
2. Las variables más predictivas según análisis SHAP fueron fracción de eyección, edad, filtrado glomerular, frecuencia cardíaca, presión arterial diastólica, beta-bloqueadores, glicemia, creatinina, presión arterial sistólica y diabetes mellitus, consistentes con el conocimiento fisiopatológico.
3. Los modelos mostraron buena calibración (Brier Score = 0,036 para el extendido y 0,096 para el reducido) y utilidad clínica demostrada mediante Decision Curve Analysis.
4. El análisis SHAP reveló relaciones no lineales e interacciones entre variables que los modelos tradicionales basados en regresión no capturan.
5. **Validación sin fuga de datos:** Se identificaron y excluyeron variables que podrían introducir fuga de datos parcial (`comp_*`, `aminas`, `reperfusion_*`, `tiempo_puerta_aguja`, CK tardío). El modelo reentrenado mantuvo un AUROC de 0,896 (IC 95 %: 0,843–0,939), confirmando que el rendimiento predictivo **no depende de variables con potencial sesgo**. Las variables más importantes en este modelo validado (edad, fracción de eyección, glicemia, índice Killip, presión arterial diastólica) son todas clínicamente legítimas y disponibles al ingreso del paciente.

### 9.2 Comparación con la Literatura

9.2.1 Rendimiento Comparativo

**Tabla 51.** Comparación con estudios previos de ML en predicción de mortalidad por IAM

Estudio	N	Modelo	AUROC	Validación
Zhu et al. 2024	20.000	XGBoost	0,93	Externa
Oliveira et al. 2023	5.000	Ensemble	0,89	Interna
Wang et al. 2022	8.500	Random Forest	0,87	Temporal
<b>Presente estudio (Reducido)</b>	3.112	XGBoost	0,90	Bootstrap/Jackknife
<b>Presente estudio (Extendido)</b>	3.112	XGBoost	0,94	Bootstrap/Jackknife

El rendimiento de nuestros modelos es comparable al de estudios previos en poblaciones internacionales, a pesar del menor tamaño muestral. El modelo extendido alcanza un AUROC de 0,938, situándose entre los mejores resultados publicados. Esta observación sugiere que la calidad y completitud de las variables clínicas del registro RECUIMA compensa parcialmente la limitación del tamaño muestral.

Las diferencias en rendimiento respecto a otros estudios pueden explicarse por varios factores: (1) la homogeneidad de la población cubana reduce la variabilidad no explicada; (2) el registro RECUIMA incluye variables detalladas del manejo terapéutico que enriquecen el modelo; (3) la tasa de mortalidad del 8,8 % proporciona suficientes eventos para el entrenamiento robusto de los modelos.

9.2.2 Variables Predictoras

Las variables identificadas como más predictivas por el análisis SHAP coinciden ampliamente con los factores incluidos en escalas tradicionales como GRACE y TIMI. La fracción de eyección, principal predictor identificado, no se incluye directamente en GRACE pero sí en otras herramientas pronósticas como el índice CADILLAC. Resulta destacable el papel del filtrado glomerular y la creatinina como predictores independientes, reflejando la importancia del síndrome cardiorenal en el pronóstico del IAM.

La identificación del uso de betabloqueadores como factor protector es consistente con las guías de práctica clínica que recomiendan su uso temprano en pacientes sin contraindicaciones. Este hallazgo sugiere que los modelos de ML pueden capturar tanto factores de riesgo como efectos beneficiosos del tratamiento adecuado.

Una fortaleza del enfoque de ML es la capacidad de identificar relaciones no lineales. Por ejemplo, el efecto de la presión arterial sistólica sigue un patrón en J, donde tanto la hipotensión severa (indicativa de shock) como valores muy elevados se asocian con mayor riesgo, mientras que valores normales confieren relativa protección.

## 9.3 Fortalezas del Estudio

1. **Metodología rigurosa:** El estudio siguió las directrices TRIPOD+AI para el desarrollo y reporte de modelos predictivos clínicos. Se implementó una separación estricta de conjuntos train/validation/test (60%/20%/20%), validación cruzada estratificada de 5 pliegues, y evaluación con múltiples métricas complementarias (discriminación, calibración, utilidad clínica).
2. **Análisis de explicabilidad:** El uso de técnicas SHAP permite interpretar las predicciones del modelo, facilitando la confianza clínica y la identificación de factores de riesgo modificables. Las explicaciones son consistentes con el conocimiento fisiopatológico establecido.
3. **Comparación con benchmarks establecidos:** La comparación directa con el score GRACE proporciona contexto clínico relevante y demuestra el valor añadido de los modelos de ML sobre las herramientas tradicionales de estratificación de riesgo.
4. **Evaluación de utilidad clínica:** El análisis de curva de decisión va más allá de métricas estadísticas para evaluar el beneficio práctico del modelo en el rango de umbrales de decisión clínicamente relevantes.
5. **Reproducibilidad:** El código fuente, pipelines de preprocesamiento y modelos entrenados están documentados y disponibles para replicación, facilitando la validación externa y extensión del trabajo.

## 9.4 Limitaciones del Estudio

### 9.4.1 Limitaciones Metodológicas

1. **Diseño retrospectivo:** Los datos fueron recolectados retrospectivamente a partir del registro RECUIMA, lo que puede introducir sesgos de selección y de información. Sin embargo, el registro sigue protocolos estandarizados de recolección que minimizan este riesgo.
2. **Cohorte nacional única:** El modelo fue desarrollado con datos de pacientes cubanos, lo que puede limitar la generalización directa a otras poblaciones con diferentes características demográficas, patrones de práctica clínica o acceso a tratamientos. Se requiere validación en cohortes externas.
3. **Validación interna:** La validación se realizó mediante división interna de datos con técnicas de Bootstrap y Jackknife. Aunque estas técnicas proporcionan esti-

maciones robustas, la validación externa en cohortes independientes es necesaria para confirmar la generalizabilidad.

4. **Datos faltantes:** Algunas variables presentaron datos faltantes que fueron manejados mediante exclusión de variables con más del 50 % de datos ausentes y estrategias de imputación. Estas decisiones metodológicas pueden introducir sesgos.

#### 9.4.2 Limitaciones de los Datos

1. **Periodo temporal:** Los datos corresponden al período 2019–2023, y los patrones de práctica clínica pueden evolucionar, especialmente tras la pandemia de COVID-19 que afectó el manejo de pacientes cardiovasculares.
2. **Variables no disponibles:** Algunas variables potencialmente relevantes como troponinas de alta sensibilidad, péptidos natriuréticos, y datos de imagen cardíaca avanzada no estaban disponibles de forma sistemática en el dataset.
3. **Definición del outcome:** La mortalidad intrahospitalaria es un endpoint bien definido pero no captura eventos post-alta (mortalidad a 30 días o 1 año) que pueden ser clínicamente relevantes para la toma de decisiones a largo plazo.

#### 9.4.3 Limitaciones del Modelo

1. **Interpretabilidad vs. rendimiento:** Aunque las técnicas SHAP mejoran sustancialmente la interpretabilidad, los modelos de gradient boosting siguen siendo más complejos que la regresión logística tradicional, lo que puede dificultar su aceptación en algunos contextos clínicos.
2. **Causalidad:** El modelo identifica asociaciones predictivas, no relaciones causales. La intervención sobre variables identificadas como factores de riesgo no garantiza necesariamente mejora en outcomes, aunque puede guiar hipótesis para investigación futura.
3. **Drift temporal:** El rendimiento del modelo puede degradarse con el tiempo si cambian las características de la población, las prácticas clínicas o los tratamientos disponibles. Se recomienda monitorización continua del rendimiento en implementación real.

## 9.5 Implicaciones Clínicas



### 9.5.1 Potencial de Implementación

El modelo desarrollado presenta características favorables para su implementación clínica. Los principales escenarios de uso incluyen: (1) triaje inicial en urgencias para identificar pacientes de alto riesgo que requieren monitorización intensiva; (2) apoyo a la decisión sobre nivel de cuidados (UCI vs. UCO vs. planta); y (3) comunicación estructurada del pronóstico al equipo clínico.

La integración con los flujos de trabajo existentes es factible dado que las variables requeridas por el modelo reducido (10 variables) están disponibles rutinariamente en la evaluación inicial del paciente con IAM. El dashboard desarrollado proporciona una interfaz intuitiva que no requiere conocimientos técnicos avanzados para su uso.

### 9.5.2 Beneficios Potenciales

1. **Identificación temprana de alto riesgo:** El modelo permite identificar en las primeras horas pacientes que requieren monitorización intensiva o intervenciones agresivas tempranas, potencialmente mejorando la supervivencia.
2. **Optimización de recursos:** La estratificación objetiva de riesgo puede guiar la asignación eficiente de camas en UCI/UCO, especialmente relevante en contextos con recursos limitados.
3. **Comunicación con pacientes/familias:** Las explicaciones SHAP proporcionan información pronóstica objetiva y comprensible que facilita la toma de decisiones compartida y el consentimiento informado.
4. **Apoyo a la decisión:** El modelo complementa, sin reemplazar, el juicio clínico, proporcionando información cuantitativa que reduce la variabilidad inter-observador en la evaluación pronóstica.

### 9.5.3 Consideraciones de Implementación

La implementación exitosa del modelo requiere considerar varios factores: (1) integración técnica con sistemas de historia clínica electrónica existentes; (2) capacitación del personal clínico en la interpretación de las predicciones y explicaciones SHAP; (3) validación local del rendimiento antes de adopción rutinaria; (4) establecimiento de protocolos de monitorización continua del rendimiento; y (5) consideración de aspectos regulatorios según la normativa local sobre software médico.

## 9.6 Consideraciones Éticas

La implementación de modelos predictivos basados en inteligencia artificial en el contexto clínico plantea importantes consideraciones éticas:

- **Equidad:** El modelo debe ser evaluado en diferentes subgrupos demográficos para identificar y mitigar posibles sesgos que pudieran resultar en disparidades en la atención.
- **Transparencia:** Las explicaciones SHAP facilitan la comprensión del razonamiento del modelo, pero deben comunicarse de forma apropiada a pacientes y familias, evitando generar ansiedad innecesaria o falsas certezas.
- **Autonomía:** El modelo es una herramienta de apoyo a la decisión clínica, no un sustituto del juicio profesional ni de la voluntad del paciente. Las predicciones deben contextualizarse dentro de la situación individual de cada paciente.
- **Privacidad:** Los datos clínicos utilizados para predicción deben manejarse según las normativas de protección de datos aplicables, garantizando la confidencialidad de la información del paciente.
- **Responsabilidad:** Es necesario establecer claramente las responsabilidades cuando las predicciones del modelo influyan en decisiones clínicas, manteniendo al profesional de salud como responsable último de la atención.

## 10 Conclusiones

### 10.1 Conclusiones Principales

Este estudio demuestra que:

1. **Viabilidad:** Es factible desarrollar modelos de aprendizaje automático para predecir la mortalidad intrahospitalaria en pacientes con IAM utilizando variables clínicas y de laboratorio disponibles rutinariamente en el contexto cubano.
2. **Rendimiento superior:** Se desarrollaron dos modelos XGBoost con diferentes enfoques:
  - **Modelo reducido (10 variables):** AUROC = 0,901 (IC 95 %: 0,855–0,937), diseñado para comparación directa con la escala GRACE.
  - **Modelo extendido (57 variables):** AUROC = 0,938 (IC 95 %: 0,884–0,977), nuestra propuesta principal con rendimiento óptimo.

Ambos modelos superaron significativamente al score GRACE tradicional (AUROC = 0,820).

3. **Calibración adecuada:** Los modelos muestran buena calibración (Brier Score = 0,036 para el modelo extendido y 0,096 para el reducido), permitiendo utilizar las probabilidades predichas para la toma de decisiones clínicas.

4. **Interpretabilidad:** Las técnicas de explicabilidad (SHAP) identificaron que las variables más influyentes son fracción de eyección, edad, filtrado glomerular, frecuencia cardíaca y presión arterial diastólica, consistentes con el conocimiento fisiopatológico establecido.
5. **Utilidad clínica:** El análisis de curva de decisión demuestra beneficio neto del modelo en el rango de umbrales de probabilidad clínicamente relevantes, especialmente en la identificación de pacientes de bajo riesgo (VPN = 0,969).
6. **Robustez sin fuga de datos:** La validación del modelo excluyendo variables con potencial fuga de datos (`comp_*`, `aminas`, `reperfusion_*`, `tiempo_puerta_aguja`, `CK_tardío`) confirmó que el rendimiento predictivo (AUROC = 0,896) no depende de información sesgada. Las variables más importantes en este modelo validado son todas clínicamente legítimas y disponibles al ingreso del paciente.

## 10.2 Respuesta a los Objetivos

### 10.2.1 Objetivo General

**Objetivo:** Desarrollar y validar un modelo de aprendizaje automático para la predicción de mortalidad intrahospitalaria en pacientes con infarto agudo de miocardio.

**Conclusión:** Se cumplió el objetivo general. Se desarrollaron dos modelos XGBoost: uno reducido con AUROC de 0,901 y uno extendido con AUROC de 0,938, validados mediante Bootstrap (1000 iteraciones) y Jackknife, que superan significativamente a los modelos tradicionales de estratificación de riesgo como GRACE.

### 10.2.2 Objetivos Específicos

1. **Caracterizar el dataset:** Se analizó un dataset de 3.112 pacientes con 185 variables originales procedentes del registro RECUIMA, identificando una tasa de mortalidad del 8,8 %. Se realizó un proceso sistemático de limpieza y preprocesamiento que resultó en la selección de 57 variables para el modelo extendido y 10 variables para el modelo reducido comparable con GRACE.
2. **Realizar análisis exploratorio:** El EDA reveló diferencias significativas entre fallecidos y supervivientes en variables clave como edad, fracción de eyección, filtrado glomerular, frecuencia cardíaca y estado hemodinámico, orientando la selección de predictores y confirmando la relevancia clínica de las variables incluidas.
3. **Desarrollar y optimizar modelos:** Se evaluaron múltiples algoritmos (KNN, Regresión Logística, Árbol de Decisión, Random Forest, XGBoost, XGBoost Balanced, LightGBM), siendo XGBoost el de mejor rendimiento tras optimización de hiperparámetros mediante validación cruzada estratificada de 5 pliegues.

4. **Comparar con escalas tradicionales:** Los modelos de ML superaron significativamente al score GRACE ( $p < 0,001$ ), con el modelo extendido logrando +11,8 puntos de AUROC y el reducido +8,1 puntos, demostrando el valor añadido de los algoritmos de aprendizaje automático sobre las escalas tradicionales de estratificación de riesgo.
5. **Analizar explicabilidad:** Las técnicas SHAP revelaron que las variables más predictivas son fracción de eyección ( $|SHAP|=0,072$ ), edad (0,051), filtrado glomerular (0,048), frecuencia cardíaca (0,043) y presión arterial diastólica (0,043), con relaciones no lineales e interacciones que los modelos lineales no capturan.
6. **Desarrollar herramienta de aplicación:** Se implementó un dashboard interactivo en Streamlit que permite realizar predicciones individuales con explicaciones SHAP integradas, facilitando su potencial uso clínico como herramienta de apoyo a la decisión.

### 10.3 Contribuciones del Estudio

Las principales contribuciones de este trabajo son:

1. **Científica:** Demostración de la superioridad de modelos de aprendizaje automático sobre escalas tradicionales como GRACE para la predicción de mortalidad intrahospitalaria en pacientes con IAM en una población latinoamericana, con mejoras significativas tanto en discriminación (AUROC +8,1 a +11,8 puntos) como en calibración.
2. **Metodológica:** Aplicación rigurosa de las guías TRIPOD+AI para el desarrollo y reporte de modelos predictivos clínicos, incluyendo validación mediante técnicas robustas (Bootstrap, Jackknife), análisis de calibración, utilidad clínica (DCA) y explicabilidad (SHAP).
3. **Validación de integridad:** Identificación y exclusión de variables con potencial fuga de datos, demostrando que el modelo mantiene su rendimiento predictivo (AUROC = 0,896) utilizando únicamente predictores clínicamente válidos y disponibles al ingreso del paciente.
4. **Práctica:** Desarrollo de una herramienta de predicción interpretable implementada como dashboard interactivo, lista para validación clínica prospectiva y potencial integración en la práctica asistencial.
5. **Educativa:** Documentación completa del proceso de desarrollo de modelos de ML en salud, incluyendo código fuente y pipelines reproducibles, útil para futuros proyectos similares en el contexto cubano y latinoamericano.

## 10.4 Mensaje Clave

### Punto Clave

#### Tres hallazgos principales:

1. **Superioridad sobre escalas tradicionales:** Los modelos XGBoost superaron significativamente al score GRACE, con el modelo extendido (57 variables) alcanzando AUROC = 0,938 y el modelo reducido (10 variables) logrando AUROC = 0,901, frente a 0,820 de GRACE.
2. **Alto valor predictivo negativo:** El modelo reducido alcanzó un VPN de 0,969, permitiendo identificar de manera confiable a pacientes de bajo riesgo que podrían beneficiarse de manejo menos intensivo.
3. **Validación de integridad (sin fuga de datos):** Se demostró que el rendimiento predictivo del modelo (AUROC = 0,896) **no depende de variables con potencial sesgo**. Al excluir variables que podrían introducir fuga de datos (`comp_*`, `aminas`, `reperfusion_*`, `tiempo_puerta_aguja`, `CK tardío`), el modelo mantuvo su capacidad discriminativa. Las variables más importantes según SHAP en este modelo validado —edad, fracción de eyección, glicemia, índice Killip, presión arterial diastólica— son todas clínicamente legítimas y disponibles al momento del ingreso del paciente, garantizando la aplicabilidad real del modelo en la práctica clínica.

### Punto Clave

#### Implicación práctica:

La validación sin fuga de datos confirma que el modelo puede implementarse de forma segura en entornos clínicos reales, ya que sus predicciones se basan exclusivamente en información disponible al momento de la toma de decisiones, sin depender de variables que solo se conocen retrospectivamente o en pacientes con evolución desfavorable.

## 10.5 Recomendaciones

### 10.5.1 Para Investigadores

1. Realizar validación externa del modelo en cohortes independientes.
2. Explorar la extensión del modelo a otros outcomes (arritmias, reingresos).
3. Investigar la incorporación de datos multimodales (imágenes, señales).

4. Desarrollar modelos de predicción dinámica durante la hospitalización.

### 10.5.2 Para Clínicos

1. Considerar el modelo como complemento, no sustituto, del juicio clínico.
2. Utilizar las explicaciones SHAP para identificar factores de riesgo modificables.
3. Validar localmente el modelo antes de su implementación rutinaria.
4. Participar en estudios de evaluación de impacto clínico.

### 10.5.3 Para Gestores de Salud

1. Evaluar la integración de herramientas de ML en sistemas de información hospitalarios.
2. Considerar la inversión en infraestructura de datos para habilitar el uso de IA clínica.
3. Establecer procesos de gobernanza para la implementación de algoritmos predictivos.
4. Facilitar la colaboración multicéntrica para validación y mejora de modelos.

## 10.6 Reflexión Final

Este estudio representa un paso significativo hacia la integración de la inteligencia artificial en la cardiología de urgencias en el contexto cubano. Los resultados demuestran que es posible desarrollar modelos predictivos de alto rendimiento utilizando datos clínicos rutinarios del registro RECUIMA, superando las limitaciones de las escalas internacionales desarrolladas en poblaciones diferentes.

Aunque los resultados son prometedores, el camino hacia la implementación clínica requiere validación rigurosa en cohortes independientes, colaboración multidisciplinaria entre informáticos, estadísticos y clínicos, y un compromiso continuo con la transparencia y la ética en el desarrollo de algoritmos de salud.

La predicción de riesgo no es un fin en sí mismo, sino una herramienta para mejorar la atención al paciente. El verdadero valor de estos modelos se materializará cuando contribuyan a salvar vidas mediante una mejor identificación de pacientes vulnerables, una asignación más eficiente de recursos sanitarios, y una comunicación más efectiva del pronóstico entre el equipo de salud, los pacientes y sus familias.

## Referencias Bibliográficas

- Antman, E. M., Cohen, M., Bernink, P. J. L. M., McCabe, C. H., Horacek, T., Papuchis, G., Mautner, B., Corbalan, R., Radley, D., & Braunwald, E. (2000). The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. *JAMA*, 284(7), 835-842. <https://doi.org/10.1001/jama.284.7.835>
- Fox, K. A. A., et al. (2006). Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). *BMJ*, 333(7578), 1091. <https://doi.org/10.1136/bmj.38985.646481.55>
- Morrow, D. A., Antman, E. M., Charlesworth, A., Cairns, R., Murphy, S. A., de Lemos, J. A., Giugliano, R. P., McCabe, C. H., & Braunwald, E. (2013). The ACTION Registry-GWTG risk score predicts mortality in acute myocardial infarction. *American Heart Journal*, 166(4), 607-615. <https://doi.org/10.1016/j.ahj.2013.06.023>
- Oliveira, M., Silva, R., Santos, A., Costa, P., & Ferreira, L. (2023). Machine learning prediction of mortality in acute myocardial infarction. *BMC Medical Informatics and Decision Making*, 23(1), 18. <https://doi.org/10.1186/s12911-023-02118-4>
- Pinto, F. J., Ferreira, J., Timoteo, A. T., Gil, V., et al. (2015). Development and validation of the ProACS risk score for predicting mortality in acute coronary syndromes. *Revista Portuguesa de Cardiologia*, 34(12), 771-781. <https://doi.org/10.1016/j.repc.2015.07.003>
- Steyerberg, E. W. (2019). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating* (2nd). Springer Nature. <https://doi.org/10.1007/978-3-030-16399-0>
- Wang, H., Li, Y., Zhang, C., Huang, J., & Liu, T. (2022). Integration of inflammatory and renal biomarkers to predict mortality in acute myocardial infarction using machine learning. *Frontiers in Cardiovascular Medicine*, 9, 948112. <https://doi.org/10.3389/fcvm.2022.948112>
- Zhu, X., Wang, L., Zhang, Y., Chen, W., Liu, M., et al. (2024). Machine learning prediction of in-hospital mortality for acute myocardial infarction: A multicenter study. *European Heart Journal – Digital Health*, 5(2), 148-159. <https://doi.org/10.1093/ehjdh/ztad074>