

Modelo predictivo basado en Random Forest para eventos duros en insuficiencia cardíaca: validación en una cohorte comunitaria real de 20 años con análisis diferenciado por sexo y espectro de FEVI

Marianela Carolina Castillo Nuñez(1), María Carmen Durán Torralba(1), Gustavo Cortez Quiroga(1) y Carlos Millán Rodríguez(1) de (1)Servicio de Cardiología. Hospital Alto Guadalquivir, Andújar (Jaén) - España

Introducción: Los scores de riesgo clásicos en insuficiencia cardíaca (IC) como el MAGGIC presentan capacidad discriminatoria limitada (AUC ~0,70–0,75) con escasa granularidad por sexo y espectro de fracción de eyección (FEVI). Los algoritmos de Machine Learning (ML) permiten integrar múltiples variables clínicas, ecocardiográficas, analíticas y farmacológicas para mejorar la predicción de eventos duros.

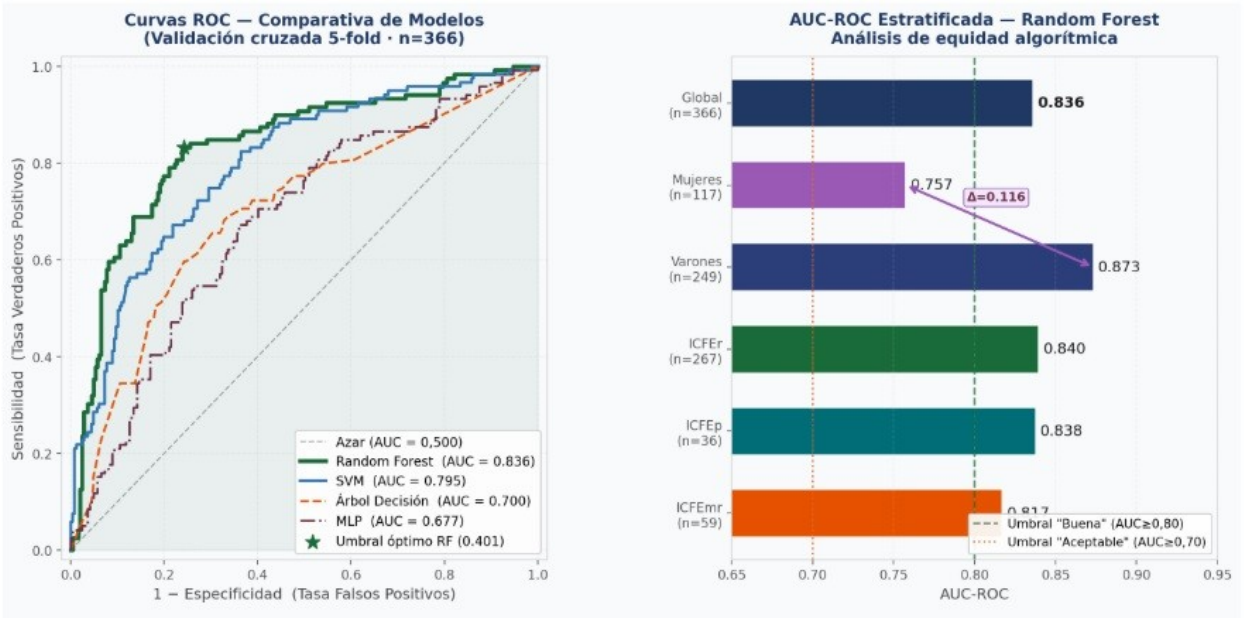
OBJETIVO: Desarrollar y validar un modelo de Random Forest (RF) para el evento duro compuesto (muerte o 1.ª hospitalización por IC) en una cohorte real de 20 años, con análisis estratificado por sexo y FEVI.

Métodos: Estudio observacional retrospectivo-prospectivo. Unidad IC Comunitaria (UICC-Andújar), Hospital Alto Guadalquivir, Andújar (Enero 2006–Abril 2026). N=472 pacientes, ?154 (32.6%); 243 variables originales. Población analítica: 366 pacientes con outcome completamente conocido en ambas variables (106 excluidos por datos parciales). Selección de features: missing ?30% ? 83 variables en 9 dominios. Preprocesamiento: imputación por mediana, StandardScaler, class_weight='balance'. Algoritmos: RF (Random Forest), SVM (Máquina de Vector de Soportes), Árbol de Decisión (CART), Perceptrón Multicapa (MLP). Validación cruzada estratificada 5-fold; umbral óptimo: índice de Youden. Métricas principales: AUC-ROC, sensibilidad, especificidad y VPN. Análisis univariante con tests paramétricos y no paramétricos según distribución.

Resultados: Evento compuesto: 119/366 (32,5%). Sin diferencias significativas por sexo (mujeres 29,9% vs varones 33,7%) ni FEVI (ICFEr 31,1%, ICFEp 38,9%, ICFEmr 32,2%). Análisis univariante: 30 predictores significativos (p<0,05); principales: Edad, NYHA, test 6 minutos, hemoglobina, filtrado glomerular y Creatinina basal. El RF fue el mejor modelo: AUC=0,836; sensibilidad 83,2%, especificidad 75,7%, VPP 62,3%, VPN 90,3%, F1=0,712, Brier=0,165. Análisis de equidad: AUC mujeres 0,757 vs varones 0,873 (?=0,116); AUC ICFEr 0,840, ICFEp 0,838, ICFEmr 0,817. Top predictores RF: creatinina 12m (4,92%), albúmina basal (4,74%), test 6 minutos basal (4,19%), colesterol total 12m (3,95%), IMC 12m (3,60%), Sodio 12m (3,44%).

Conclusiones: El RF alcanza AUC=0,836 superando los scores clásicos. La brecha de rendimiento por sexo (?=0,116) justifica el desarrollo de modelos diferenciados por sexo como siguiente fase del proyecto. El VPN del

90,3% ofrece alta seguridad para descartar riesgo bajo.



Métricas de desempeño de los modelos de Machine Learning evaluados (n=366)

MODELO	AUC-ROC	SENS.	ESP.	VPP	VPN	F1	BRIER
RF	**0,836**	**83,2%**	**75,7%**	**62,3%**	**90,3%**	**0,712**	**0,165**
SVM	0,795	82,4%	63,6%	52,1%	88,2%	0,638	0,165
CART	0,700	59,7%	75,7%	54,2%	79,6%	0,568	0,244
MLP	0,677	67,2%	64%	47,3%	80,2%	0,556	0,202

ACU(Área bajo la curva de la característica operativa del receptor);RF (Random Forest); SVM (Máquina de Vector de Soportes); Árbol de Decisión (CART); Perceptrón Multicapa (MLP); SENS: sensibilidad; ESP: especificidad; VPP: valor predictivo positivo; VPN: valor predictivo negativo; F1: F1-score; BRIER: Brier Score.