

Random forest frente a regresión logística q-BH para predecir shock cardiogénico en síndrome coronario agudo

Álvaro Jara Díaz(1), Laura Teresa Botella Pascual(1), Sergi Felis Badia(1), M. Isabel Hernández Martín(1), Gemma Mateus Porta(1), Luis Enrique Martín Alfaro(1), Esther Papiol Morales(1), Arnau Cabello Ramírez(1), Mireia Martínez Rubió(1), Alba Orellana Morote(1), Alba Bermúdez Jiménez(1), Tania Ramírez Martínez(1) y Juan Manuel Casanova Sandoval(1) de (1)Servicio de Cardiología. Hospital Universitari Arnau de Vilanova, Lleida - España

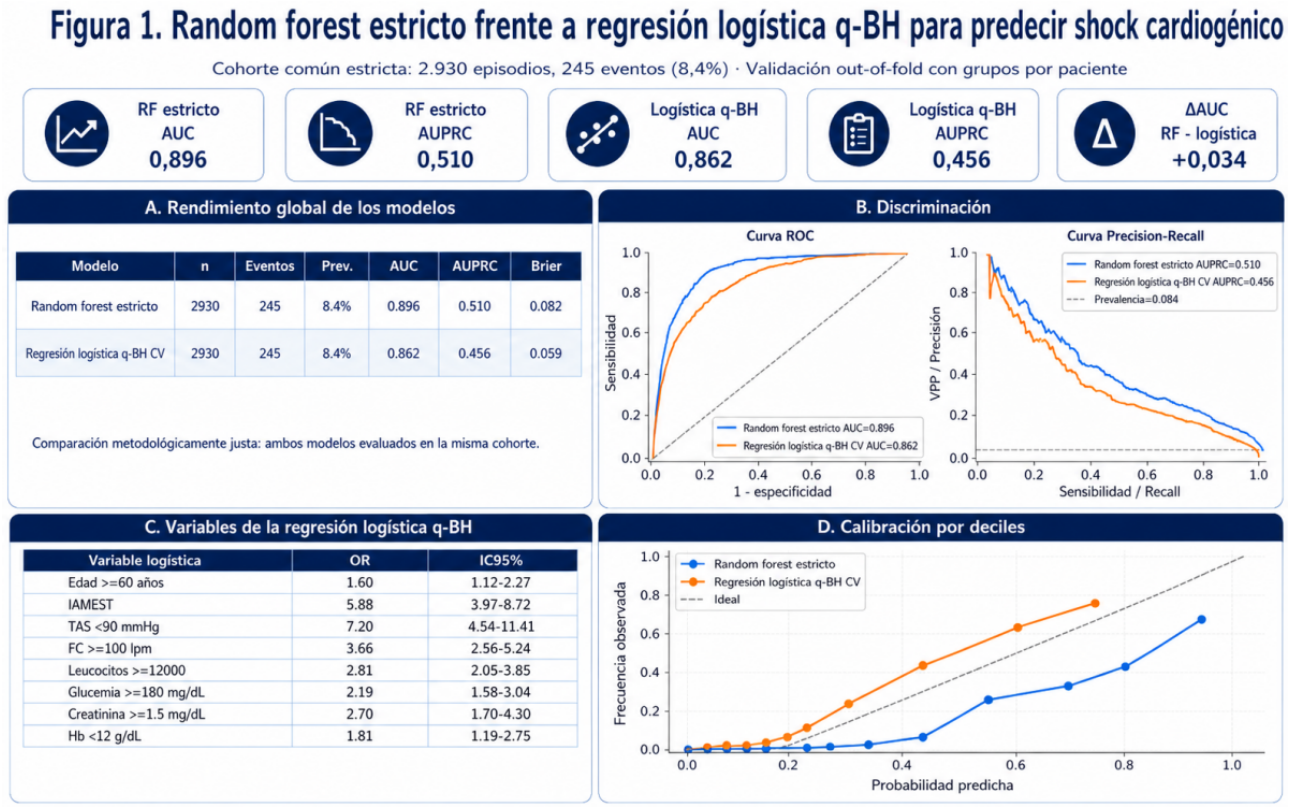
Introducción: El shock cardiogénico en el síndrome coronario agudo es una complicación de alta mortalidad, por lo que la estratificación precoz del riesgo resulta clínicamente relevante. La regresión logística aporta interpretabilidad y estimaciones ajustadas, mientras que algoritmos de machine learning como random forest pueden capturar relaciones no lineales e interacciones. El objetivo fue comparar, en una misma cohorte, el rendimiento de un random forest estricto frente a una regresión logística q-BH validada por cross-validation para predecir shock cardiogénico corregido.

Métodos: Estudio observacional retrospectivo de pacientes ingresados por síndrome coronario agudo entre 2016 y 2025. El análisis se realizó en una cohorte común estricta de 2.930 episodios, con 245 eventos de shock cardiogénico, prevalencia 8,4%. El random forest fue reentrenado exclusivamente en esta cohorte y evaluado mediante predicciones out-of-fold agrupadas por paciente. La regresión logística q-BH se evaluó en la misma cohorte mediante validación cruzada. Se compararon AUC, AUPRC y Brier score. Como referencia clínica interpretable, el modelo logístico incluyó edad ?60 años, IAMEST, TAS <90 mmHg, FC ?100 lpm, leucocitos ?12.000/mm³, glucemia ?180 mg/dL, creatinina ?1,5 mg/dL y Hb <12 g/dL.

Resultados: El random forest estricto mostró mejor discriminación que la regresión logística q-BH CV: AUC 0,896 frente a 0,862 y AUPRC 0,510 frente a 0,456. La diferencia absoluta fue +0,034 en AUC y +0,054 en AUPRC a favor del random forest. Sin embargo, la regresión logística presentó mejor Brier score, 0,059 frente a 0,082, sugiriendo mejor calibración global de las probabilidades. En el modelo logístico, las asociaciones más intensas correspondieron a TAS <90 mmHg, OR 7,20; IC95% 4,54–11,41, IAMEST, OR 5,88; IC95% 3,97–8,72, y FC ?100 lpm, OR 3,66; IC95% 2,56–5,24. También se asociaron leucocitos ?12.000/mm³, OR 2,81; creatinina ?1,5 mg/dL, OR 2,70; glucemia ?180 mg/dL, OR 2,19; Hb <12 g/dL, OR 1,81; y edad ?60 años, OR 1,60.

Conclusiones: En esta cohorte retrospectiva de síndrome coronario agudo, el random forest superó a la regresión logística q-BH en discriminación para predecir shock cardiogénico, mientras que la logística mostró mejor Brier score y mayor interpretabilidad clínica. Ambos enfoques pueden considerarse complementarios: random forest para optimizar el ranking predictivo y regresión logística para comunicación clínica e interpretación fisiopatológica. Estos resultados deben considerarse exploratorios y requieren validación temporal o externa antes de su

implementación clínica.



Variables de la regresión logística q-BH

VARIABLE LOGÍSTICA	OR	IC 95%
Edad >= 60 años	1.60	1.12-2.27
IAMEST	5.88	3.97-8.72
TAS < 90 mmHg	7.20	4.54-11.41
FC > 100 lpm	3.66	2.56-5.24
Leucocitos >= 12000	2.81	2.05-3.85
Glucemia >= 180 mg/dL	2.19	1.58-3.04
Creatinina >= 1.5mg/dL	2.70	1.70-4.30
Hb < 12 g/dL	1.81	1.19-2.75