Avance online de artículo en prensa

ORIGINAL

Modelos predictivos basados en aprendizaje automático para apoyar la selección del tratamiento empírico en pacientes con bacteriemia atendidos en el servicio de urgencias

Álvaro Monje[1-3], Laura Escolà-Vergé[2,4-6], Celso Soares[5,7,8], Mireia Puig[2,9], Sergio Herrera[9], Jesús Ruiz[1,2]

Objetivo. Desarrollar y comparar modelos predictivos para estimar la cobertura empírica inadecuada con ceftriaxona en pacientes con bacteriemia atendidos en el servicio de urgencias (SU) a partir de variables disponibles en la valoración inicial.

Método. Cohorte retrospectiva unicéntrica de pacientes adultos consecutivos con bacteriemia. Se desarrollaron siete modelos (regresión logística, XGBoost, CatBoost, LightGBM, Random Forest, Support Vector Machine y Naive Bayes) con variables clínicas rutinarias, optimización de hiperparámetros, ponderación de clases y ajuste del umbral para priorizar la sensibilidad. El rendimiento se evaluó en validación interna (n = 336) mediante el área bajo la curva de la característica operativa del receptor (ABC COR), sensibilidad, especificidad, coeficiente de correlación de Matthews y puntuación de Brier, con IC 95% por remuestreo. Se evaluaron la utilidad clínica mediante *Decision Curve Analysis* (DCA) y la interpretabilidad mediante Explicaciones Aditivas de Shapley (SHAP).

Resultados. Se incluyeron 1.678 pacientes, 678 (40,4%) con microorganismos no cubiertos adecuadamente por ceftriaxona. Todos los modelos mostraron una capacidad discriminativa moderada, sin diferencias relevantes entre modelos (ABC COR 0,716-0,732; IC 95% solapados). La regresión logística alcanzó un ABC COR de 0,731 (IC 95% 0,677-0,786), una sensibilidad del 70,6% y una especificidad del 62,0%. Los principales predictores (SHAP) fueron la infección relacionada con la asistencia sanitaria, el foco de catéter y la colonización previa por microorganismo multirresistente. El DCA mostró beneficio neto frente a tratar a todos o a ningún paciente, sin diferencias relevantes entre modelos.

Conclusiones. Las variables disponibles en la valoración inicial en el SU permiten estimar precozmente la cobertura empírica inadecuada con ceftriaxona. Como no se detectan diferencias relevantes de rendimiento, la regresión logística ofrece ventajas de interpretabilidad e implementación. Se trata de una propuesta pendiente de validación externa antes de cualquier uso asistencial para su integración en herramientas de apoyo a la decisión clínica.

Palabras clave: Ceftriaxona. Bacteriemia. Aprendizaje automático. Servicio de Urgencias. Adecuación tratamiento empírico.

Machine learning-based predictive models to support empirical treatment selection in patients with bacteremia attended in the emergency department

Objective. To develop and compare predictive models for estimating inadequate empirical ceftriaxone coverage in patients with bacteremia attended in the emergency department (ED), using variables available at the initial assessment.

Methods. Single-center retrospective cohort of consecutive adult patients with bacteremia. Seven models were developed (logistic regression, XGBoost, CatBoost, LightGBM, Random Forest, Support Vector Machine, and Naive Bayes) using routine clinical variables, hyperparameter optimization, class weighting, and threshold adjustment to prioritize sensitivity. Performance was assessed by internal validation (n = 336) using AUROC, sensitivity, specificity, Matthews correlation coefficient, and Brier score, with 95%CI obtained by bootstrap. Clinical utility was assessed using Decision Curve Analysis (DCA), and interpretability using Shapley Additive Explanations (SHAP).

Results. A total of 1,678 patients were included, 678 (40.4%) with microorganisms not adequately covered by ceftriaxone. All models showed moderate discriminatory ability, with no relevant differences between models (AUROC, 0.716–0.732; overlapping 95%CI). Logistic regression achieved an AUROC of 0.731 (95%CI, 0.677–0.786), a sensitivity rate of 70.6%, and a specificity rate of 62.0%. The main predictors (SHAP) were health care-associated infection, catheter-related source, and previous colonization by a multidrug-resistant microorganism. DCA showed net benefit compared with treating all patients or none, with no relevant differences across models.

Conclusions. Variables available at the initial ED assessment allow early estimation of inadequate empirical ceftriaxone coverage. When no relevant differences in performance are detected, logistic regression offers advantages in

Filiación de los autores:
[1]Servicio de Farmacia, Hospital de la Santa Creu i Sant Pau, Barcelona, España.
[2]Departament de Medicina, Universitat Autònoma de Barcelona, Unitat Docent Vall d'Hebron, Barcelona, España.
[3]Centro Médico Teknon. Quirónsalud, Barcelona, España.
[4]Servicio de Enfermedades Infecciosas, Hospital de la Santa Creu i Sant Pau, Barcelona, España.
[5]Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, España.
[6]Centro de Investigación Biomédica en Red de Enfermedades Infecciosas (CIBERINFEC), Instituto de Salud Carlos III, Madrid, España.
[7]Servicio de Microbiología, Hospital de la Santa Creu i Sant Pau, Barcelona, España.
[8]Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Barcelona, España.
[9]Servicio de Urgencias, Hospital de la Santa Creu i Sant Pau, Barcelona, España

Contribución de los autores:
Todos los autores han confirmado su autoría en el documento de responsabilidades del autor, acuerdo de publicación y cesión de derechos a EMERGENCIAS.

Autor para correspondencia:
Álvaro Monje López.
Servicio de Farmacia.
Hospital de la Santa Creu i Sant Pau.
Carrer de Sant Quintí 89. 08025 Barcelona, España.

Correo electrónico:
alvaro.monjelopez96@gmail.com

Información del artículo:
Recibido: 24-7-2026
Aceptado: 21-8-2026
Online: 13-11-2026

Editor responsable:
Juan González del Castillo

DOI:
XXXX

Avance online de artículo en prensa

Avance online de artículo en prensa

interpretability and implementation. This proposal requires external validation before any clinical use and before integration into clinical decision-support tools.

Keywords: Ceftriaxone. Bacteremia. Machine learning. Emergency department. Appropriateness of empirical treatment.

DOI: XXXXX

Introducción

La resistencia antimicrobiana constituye uno de los principales problemas de salud pública, responsable de millones de muertes anuales y de una reducción significativa de las opciones terapéuticas disponibles[1]. La infección por microorganismos multirresistentes (MMR) se asocia con mayor mortalidad, estancias hospitalarias prolongadas y aumento de los costes sanitarios[2]. Entre ellos, los bacilos gramnegativos productores de β-lactamasas de espectro extendido (BLEE) representan una causa frecuente de bacteriemia y sepsis en los servicios de urgencias (SU)[3].

La ceftriaxona es uno de los antibióticos más utilizados en los SU para el tratamiento empírico de múltiples síndromes infecciosos, pero el aumento de la prevalencia a nivel comunitario de *Enterobacterales* productores de BLEE y AmpC (20-25% según el registro VINCat) ha reducido su eficacia[4,5]. Dado que los métodos convencionales de sensibilidad requieren entre 24 y 72 horas, la elección del tratamiento empírico debe basarse inicialmente en la valoración clínica y la epidemiología local[6].

La bacteriemia es la causa principal de sepsis y *shock* séptico, y se asocia a mayor mortalidad cuando el tratamiento empírico inicial no es activo[7]. Las guías de la Surviving Sepsis Campaign 2026 priorizan la administración temprana de antibióticos activos y recomiendan reservar la cobertura frente a MMR para pacientes con factores de riesgo específicos[8]. No obstante, en la práctica habitual de los SU, esta estrategia puede favorecer el uso innecesario de carbapenémicos, ya que identificar dichos factores de riesgo resulta complejo dada la elevada presión asistencial, el corto intervalo de tiempo disponible y la frecuente presencia de pacientes con alteración del nivel de conciencia. Además, la prescripción iniciada en el SU suele mantenerse sin reevaluación durante las primeras 48-72 horas e incluso durante transiciones asistenciales, por lo que una elección empírica inadecuada puede conducir tanto a cobertura insuficiente como a uso innecesario de antibióticos de amplio espectro.

El principal reto en este contexto consiste en identificar precozmente a los pacientes con mayor probabilidad de presentar microorganismos no cubiertos adecuadamente por ceftriaxona. Diversos factores clínicos se han asociado con resistencia antimicrobiana, pero su integración en modelos predictivos en los SU es limitada, ya que la mayoría se han desarrollado en pacientes hospitalizados o en unidades de cuidados intensivos[9-11]. En el SU, Chiang *et al.* desarrollaron un modelo de predicción de bacteriemia por gramnegativos resistentes[12], referido a la resistencia microbiológica y no a la adecuación de un antibiótico concreto. Las revisiones sistemáticas disponibles sobre modelos predictivos en los SU se han centrado además en predecir la presencia de bacteriemia[13] o la indicación de hemocultivos ante la sospecha de sepsis[14], no la adecuación del tratamiento empírico.

Por ello, los modelos predictivos basados en aprendizaje automático han emergido como herramientas prometedoras para estimar la cobertura óptima del tratamiento empírico utilizando variables clínicas disponibles durante la atención asistencial. Sin embargo, continúa existiendo incertidumbre acerca de si los algoritmos más complejos ofrecen ventajas reales frente a modelos más simples e interpretables, como la regresión logística (RL).

El objetivo de este estudio fue desarrollar, comparar y someter a validación interna distintos modelos predictivos basados en aprendizaje automático con algoritmos simples e interpretables y métodos más complejos, para predecir la cobertura empírica inadecuada con ceftriaxona en pacientes con bacteriemia atendidos en los SU, y evaluar su potencial utilidad como herramienta de apoyo a la prescripción. La hipótesis de partida fue que las variables disponibles durante la valoración inicial permiten estimar dicha probabilidad con una capacidad discriminativa clínicamente útil, y que los algoritmos de mayor complejidad no son superiores frente a RL.

Método

Diseño del estudio

Se realizó un estudio observacional retrospectivo de desarrollo y validación de modelos predictivos en pacientes adultos (≥ 18 años) con bacteriemia atendidos en el SU de un hospital universitario terciario con aproximadamente 120.000 consultas anuales al SU. Se incluyeron consecutivamente todos los episodios registrados entre noviembre de 2021 y diciembre de 2024 con hemocultivos positivos y aislamiento de microorganismos patogénicos.

Se excluyeron los aislamientos sugestivos de contaminación (estafilococos coagulasa-negativos, *Corynebacterium* spp., *Bacillus* spp. no *anthracis* y *Cutibacterium acnes*) aislados en un único frasco sin correlación clínica, así como las candidemias (Figura 1). Los datos se obtuvieron por revisión manual de la historia clínica electrónica y de la base de datos de microbiología, siguiendo la guía TRIPOD-AI[15].

Responsabilidades éticas

El estudio fue aprobado por el Comité de Ética de Investigación del centro (Ref. IIBSP-MLA-2025-49), de acuerdo con la Declaración de Helsinki. Debido a su

Avance online de artículo en prensa

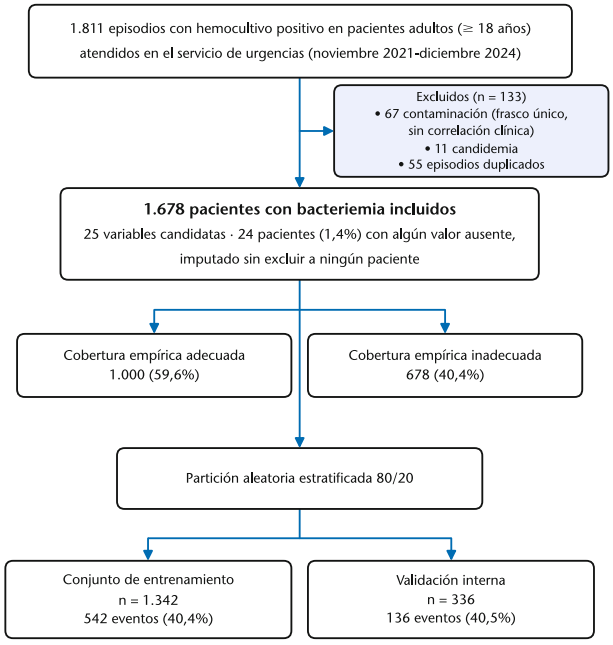


Figura 1. Diagrama de flujo del estudio: selección de la cohorte, criterios de exclusión, datos disponibles para el análisis y distribución entre los conjuntos de entrenamiento y validación interna.

naturaleza retrospectiva y al uso de datos anonimizados, se eximió el consentimiento informado individual.

Los datos que sustentan este estudio no están disponibles públicamente por contener información clínica de pacientes, pero pueden solicitarse razonadamente al autor de correspondencia. El código empleado en el análisis está disponible bajo petición.

Variables

La variable dependiente fue la adecuación de la cobertura empírica con ceftriaxona, considerada inadecuada frente a bacilos gramnegativos no fermentadores, *Staphylococcus aureus* (sensible o resistente a meticilina), estafilococos coagulasa negativos meticilín-resistentes, *Enterococcus* spp., *Streptococcus* grupo *viridans* con resistencia confirmada, *Enterobacterales* productores de BLEE o AmpC, y anaerobios. Dieciocho especies infrecuentes se clasificaron por defecto como inadecuadas ante la ausencia de puntos de corte establecidos, siguiendo las reglas de experto EUCAST[16].

Se seleccionaron 25 variables clínicas y epidemiológicas disponibles en la valoración inicial (Tabla 1) por relevancia clínica y disponibilidad sistemática, sin filtrado univariante, con una razón de eventos por variable de 21,7 en la cohorte de entrenamiento y 27,1 en la completa. La edad se incluyó como variable continua.

Todas las variables corresponden a información disponible antes del resultado microbiológico. El foco se codificó a partir de la sospecha clínica registrada en la valoración inicial en el SU, no del adjudicado tras conocer el aislamiento. El resto de datos procede de la historia clínica electrónica, accesible durante la atención del paciente.

Los valores ausentes se limitaron a las ocho comorbilidades y afectaron a 24 pacientes (1,4%); las 17 variables restantes estuvieron completas. Se imputaron mediante la mediana calculada solo sobre el conjunto de entrenamiento y aplicada al de validación, sin excluir a ningún paciente. Los contrastes de la Tabla 1 se realizaron sobre casos completos.

Las variables continuas se describieron mediante mediana y rango intercuartil y las categóricas mediante frecuencias y porcentajes, comparadas con U de Mann-Whitney y χ^2 respectivamente, con significación en $p < 0,05$.

Desarrollo y validación de los modelos

La cohorte se dividió mediante partición aleatoria estratificada 80/20 en un conjunto de entrenamiento (n = 1.342; 542 eventos) y uno de validación interna (n = 336; 136 eventos), preservando la distribución del desenlace.

Para evitar la fuga de datos, el preprocesamiento se ajustó solo sobre el entrenamiento y fue idéntico para los siete modelos: RL, Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), XGBoost, LightGBM y CatBoost.

Dado el desbalance moderado de clases (40,4% de cobertura inadecuada), no se aplicaron técnicas de sobremuestreo (SMOTE), con el fin de preservar la escala de las probabilidades predichas.

Los hiperparámetros se optimizaron mediante Optuna[17] sobre el conjunto de entrenamiento, con ABC COR como métrica objetivo. El umbral de decisión se determinó por validación cruzada, priorizando una sensibilidad ≥ 70% y, entre los candidatos, el mayor coeficiente de correlación de Matthews (MCC). El criterio se fijó *a priori* porque el coste de un tratamiento inactivo supera al de una ampliación transitoria reevaluable a las 48-72 horas. La ponderación de clases y el ajuste del umbral desplazan la escala absoluta de probabilidades sin alterar el orden de riesgo, por lo que las métricas de clasificación son invariantes.

Evaluación del rendimiento

El rendimiento se evaluó en la cohorte de validación mediante ABC COR, sensibilidad, especificidad, valor predictivo positivo y negativo, F1, MCC y puntuación de Brier, con intervalos de confianza del 95%. La interpretabilidad se evaluó mediante SHAP18, limitada a RL y XGBoost, y la utilidad clínica mediante análisis de la curva de decisión (ACD)[19].

La calibración del modelo final se evaluó mediante su curva, pendiente e intercepto. El beneficio neto del ACD se calculó sobre probabilidades recalibradas, al depender de la escala absoluta.

Análisis estadístico

Todos los análisis se realizaron con Python 3.12 y las bibliotecas scikit-learn, XGBoost, LightGBM, CatBoost, Optuna y SHAP.

Avance online de artículo en prensa

Tabla 1. Características basales de los pacientes según la adecuación de la cobertura empírica con ceftriaxona

Variable	Ceftriaxona adecuada N = 1.000 n (%)	Ceftriaxona inadecuada N = 678 n (%)	Valo de p	Perdidos n (%)
Edad [mediana (RIC)]	77 (67-86)	78 (68-86)	0,238	0
Sexo masculino	524 (52,4)	409 (60,3)	**0,001**	0
Comorbilidades				
Diabetes mellitus	282 (28,2)	202 (29,8)	0,663	21 (1,3)
EPOC	125 (12,5)	108 (15,9)	0,069	21 (1,3)
Insuficiencia renal crónica	212 (21,2)	156 (23,0)	0,521	22 (1,3)
Insuficiencia cardiaca	176 (17,6)	133 (19,6)	0,400	21 (1,3)
Hepatopatía crónica	63 (6,3)	34 (5,0)	0,226	21 (1,3)
Neoplasia	314 (31,4)	211 (31,1)	0,672	22 (1,3)
Trasplante hematológico o de órgano sólido	17 (1,7)	22 (3,2)	**0,046**	21 (1,3)
Deterioro cognitivo	141 (14,1)	158 (23,3)	< **0,001**	22 (1,3)
Institucionalización en residencia	141 (14,1)	185 (27,3)	< **0,001**	0
Ingreso hospitalario en el año previo	401 (40,1)	374 (55,2)	< **0,001**	0
Ingreso en centro sociosanitario	42 (4,2)	85 (12,5)	< **0,001**	0
Infección relacionada con la asistencia sanitaria	571 (57,1)	544 (80,2)	< **0,001**	0
Antecedente de colonización previa por MMR (12 meses)	89 (8,9)	144 (21,2)	< **0,001**	0
Foco				
Urinario	414 (41,4)	276 (40,7)	0,777	0
Abdominal	230 (23,0)	92 (13,6)	< **0,001**	0
Respiratorio	104 (10,4)	10 (1,5)	< **0,001**	0
Piel y partes blandas	41 (4,1)	83 (12,2)	< **0,001**	0
Catéter	37 (3,7)	112 (16,5)	< **0,001**	0
Uso de antibiótico 3 meses previos				
Carbapenémicos	16 (1,6)	32 (4,7)	< **0,001**	0
β-lactámicos	99 (9,9)	90 (13,3)	**0,032**	0
Quinolonas	40 (4,0)	44 (6,5)	**0,022**	0
Cotrimoxazol	33 (3,3)	29 (4,3)	0,298	0
Fosfomicina/nitrofurantoína	30 (3,0)	29 (4,3)	0,163	0

MMR: microorganismo multirresistente; RIC: rango intercuartil; EPOC: enfermedad pulmonar obstructiva crónica.
Los porcentajes se calculan sobre el total de cada grupo. Los datos de comorbilidad no estaban disponibles en 24 pacientes: 21 sin ninguna de las ocho comorbilidades registradas, 1 sin el dato de deterioro cognitivo y 2 con valores no interpretables; los contrastes se realizaron sobre casos completos. Comparaciones mediante U de Mann-Whitney (variables continuas) y prueba de χ^2 o exacta de Fisher (categóricas). Los valores perdidos afectaron a 22 pacientes (1,3%) y se concentraron exclusivamente en las variables de comorbilidad; se imputaron mediante la mediana calculada sobre el conjunto de entrenamiento, sin excluir a ningún paciente. La edad se describe mediante mediana (RIC) y media (DE); la comparación entre grupos se realizó una sola vez (U de Mann-Whitney, p = 0,238).
Los valores en negrita denotan significación estadística (p < 0,05).

Perspectiva de sexo y género

El sexo se registró a partir de la historia clínica; no se dispuso de información sobre identidad de género ni se realizaron análisis estratificados por sexo.

Resultados

El proceso de selección de la cohorte se resume en la Figura 1. Se incluyeron 1.678 pacientes con bacteriemia, de los cuales 678 (40,4%) fueron clasificados como cobertura empírica inadecuada. Las características basales se muestran en la Tabla 1: los pacientes con cobertura inadecuada presentaron mayor prevalencia de exposición sanitaria previa y de foco relacionado con catéter o de piel y partes blandas, mientras que el foco respiratorio predominó en el grupo con cobertura adecuada (p < 0,001). El sexo masculino fue más frecuente en el grupo con cobertura inadecuada (60,3% frente a 52,4%; p = 0,001), sin diferencias en la edad.

Atendiendo al criterio que determinó la clasificación, de los 678 casos de cobertura inadecuada, 510 (75,2%) correspondieron a no sensibilidad in vitro, 95 (14,0%) a *Staphylococcus aureus* sensible a meticilina por criterio terapéutico, 42 (6,2%) a especies infrecuentes sin punto de corte establecido según EUCAST y 31 (4,6%) a anaerobios.

El rendimiento de los modelos predictivos en la cohorte de validación se resume en la Tabla 2. El ABC COR osciló entre 0,716 y 0,732, con intervalos de confianza del 95% solapados entre todos los modelos. Tras la optimización del umbral de decisión (sensibilidad ≥ 70%, máximo MCC), RL alcanzó una sensibilidad de 70,6% y una especificidad de 62,0%, sin que se detectaran diferencias relevantes respecto al resto de algoritmos.

Las curvas COR se presentan en la Figura 2a. El análisis de utilidad clínica mediante ACD (Figura 2b) mostró que todos los modelos superaron las estrategias de tratar a todos o a ningún paciente en el rango de umbrales evaluado, sin diferencias relevantes entre algoritmos, incluido el punto de operación seleccionado (0,455).

El análisis SHAP se muestra en la Figura 3. En ambos modelos, la infección relacionada con la asistencia sanitaria fue el predictor de mayor peso, seguida del foco

Avance Online de artículo en prensa

Avance online de artículo en prensa

Tabla 2. Rendimiento de los modelos predictivos para estimar la cobertura empírica inadecuada con ceftriaxona en la cohorte de validación (n = 336)

Modelo	ABC COR	Sensibilidad	Especificidad	TN	VN	VPP	VPN	F1-score	MCC	Brier	Umbral
CatBoost	0,732 [0,677-0,786]	0,749 [0,674-0,820]	0,575 [0,503-0,642]	115	34	0,546 [0,474-0,616]	0,771 [0,707-0,836]	0,631 [0,564-0,691]	0,321 [0,215-0,419]	0,206 [0,192-0,218]	0,406
RL	0,731 [0,677-0,786]	0,706 [0,629-0,777]	0,620 [0,550-0,686]	124	40	0,559 [0,481-0,636]	0,755 [0,692-0,821]	0,623 [0,553-0,686]	0,320 [0,221-0,420]	0,209 [0,191-0,226]	0,455
SVM	0,731 [0,675-0,784]	0,705 [0,631-0,780]	0,635 [0,568-0,701]	127	40	0,568 [0,493-0,641]	0,759 [0,697-0,823]	0,629 [0,559-0,690]	0,334 [0,230-0,430]	0,202 [0,181-0,222]	0,356
NB	0,726 [0,668-0,779]	0,699 [0,620-0,770]	0,640 [0,574-0,703]	128	41	0,570 [0,494-0,645]	0,757 [0,695-0,818]	0,627 [0,557-0,690]	0,333 [0,227-0,433]	0,234 [0,199-0,270]	0,208
RF	0,721 [0,665-0,773]	0,662 [0,584-0,738]	0,651 [0,584-0,714]	130	46	0,564 [0,484-0,639]	0,738 [0,677-0,801]	0,608 [0,537-0,669]	0,307 [0,199-0,406]	0,211 [0,196-0,227]	0,485
XGBoost	0,718 [0,660-0,772]	0,647 [0,570-0,724]	0,670 [0,607-0,730]	134	48	0,572 [0,490-0,648]	0,736 [0,678-0,799]	0,606 [0,539-0,667]	0,312 [0,210-0,409]	0,207 [0,190-0,221]	0,386
LightGBM	0,716 [0,657-0,769]	0,669 [0,590-0,742]	0,666 [0,600-0,728]	133	45	0,577 [0,497-0,653]	0,747 [0,689-0,810]	0,619 [0,550-0,679]	0,329 [0,220-0,426]	0,207 [0,191-0,222]	0,396

ABC COR: área bajo la curva de la característica operativa del receptor; VN: verdaderos negativos (n); FN: falsos negativos (n); VPP: valor predictivo positivo; VPN: valor predictivo negativo; F1-score: media armónica de precisión y sensibilidad; MCC: coeficiente de correlación de Matthews; Brier: puntuación de Brier; SVM: Support Vector Machine; LightGBM: Light Gradient Boosting Machine; XGBoost: Extreme Gradient Boosting; CatBoost: Categorical Boosting. Cohorte de validación interna (n = 336; 136 casos con cobertura inadecuada y 200 con cobertura adecuada). Las métricas dependientes del umbral se obtuvieron tras la optimización del umbral de decisión (sensibilidad ≥ 70% y máximo MCC). Las estimaciones puntuales y sus intervalos de confianza del 95% corresponden a la media y a los percentiles 2,5 y 97,5 de 1.000 réplicas de remuestreo sobre la cohorte de validación. TN y FN se presentan como recuentos observados en dicha cohorte (136 casos con cobertura inadecuada y 200 con cobertura adecuada). La puntuación de Brier corresponde a las probabilidades predichas sin recalibrar (véase el pie de la Figura 2).

relacionado con catéter y la colonización previa por MMR. El foco respiratorio, sin embargo, contibuyó en dirección opuesta y el sexo masculino mostró una contribución relevante en la RL.

Modelo final propuesto

El modelo final con regularización L1 seleccionó 13 de las 25 variables candidatas. Solo se mantuvieron como comorbilidad el deterioro cognitivo y, entre los antibióticos previos, las quinolonas, con un coeficiente cercano a cero.

La probabilidad individual se obtiene como p = 1/(1 + e–z), donde:

z = –1,020 + 1,289·catéter + 1,010·piel y partes blandas + 0,711·IRAS + 0,640·colonización MMR + 0,441·centro sociosanitario + 0,339·residencia + 0,319·deterioro cognitivo + 0,218·sexo masculino + 0,114·ingreso previo + 0,003·quinolonas – 0,013·edad tipificada – 0,253·abdominal – 1,208·respiratorio

Las variables binarias toman valor 1 si están presentes y 0 en caso contrario; la edad se tipifica como (edad – 74,6)/15,6. Se clasifican como de alto riesgo los pacientes con p ≥ 0,455.

La calibración del modelo mostró una pendiente compatible con la unidad (1,14; IC 95% 0,81-1,47) y un intercepto de –0,385 (IC 95% –0,62 a –0,15), indicativo

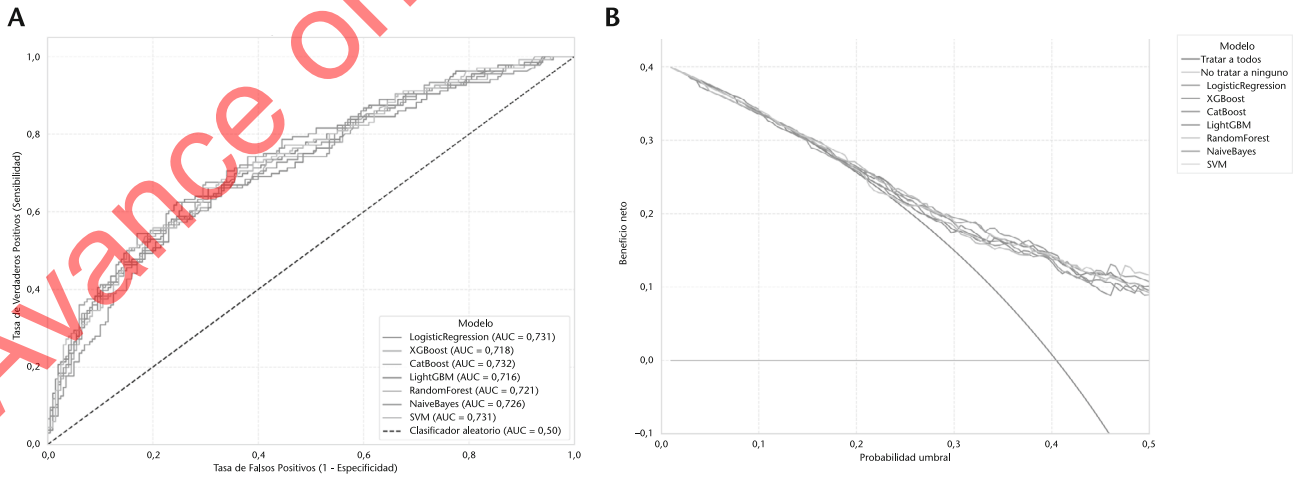


Figura 2. (A) Curvas COR (característica operativa del receptor) de los siete modelos. (B) Análisis de la curva de decisión en la cohorte de validación interna.
En el panel B *(Decision Curve Analysis)* las probabilidades predichas se recalibraron (*Platt scaling* o regresión isotónica según el modelo) para el cálculo del beneficio neto; las métricas de la Tabla 2 corresponden a las probabilidades sin recalibrar.

Avance online de artículo en prensa

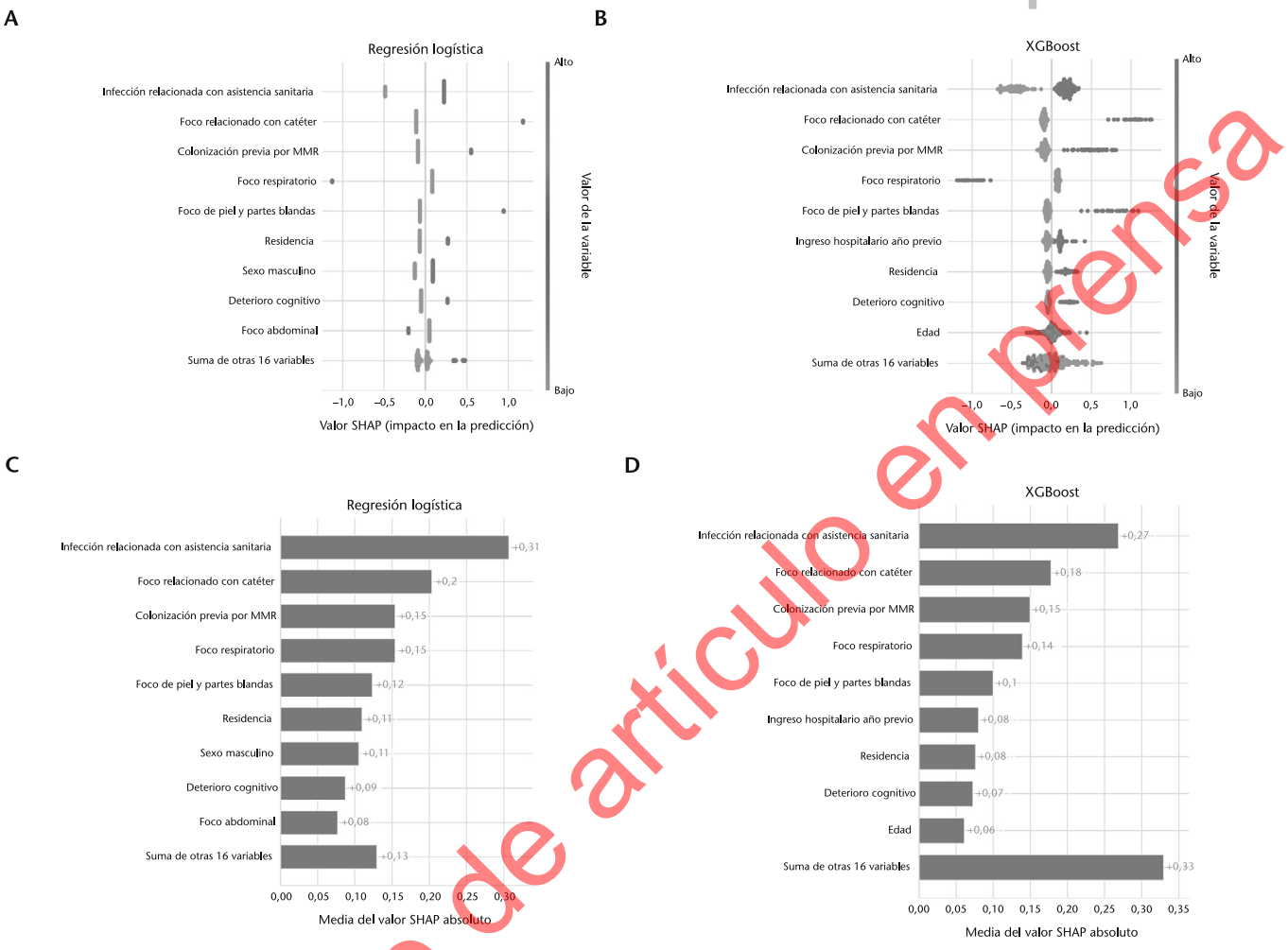


Figura 3. Interpretabilidad de los modelos mediante análisis SHAP. (A-B) Beeswarm plots y (C-D) importancia media absoluta de las variables, para la regresión logística (A, C) y XGBoost (B, D).

de una sobrestimación sistemática atribuible a la ponderación de clases; al tratarse de un desplazamiento constante en la escala logit, las métricas de la Tabla 2 no se ven afectadas, y el intercepto recalibrado (–1,406) permite estimar el riesgo absoluto (Material suplementario https://revistaemergencias.org/extra/Material-165.pdf).

La ABC COR corregida por optimismo mediante remuestreo fue de 0,748 (optimismo 0,011).

En la validación temporal (entrenamiento 2021-2023; validación 2024, n = 493, 45,0% de eventos) el modelo mantuvo su capacidad discriminativa (ABC COR 0,764; IC 95% 0,722-0,806) pese al aumento de la prevalencia respecto al periodo previo (38,5%; p = 0,015), con una pendiente de calibración de 1,375 (IC 95% 1,081-1,669) y un intercepto de –0,190 (IC 95% –0,380 a 0,000) (Material suplementario https://revistaemergencias.org/extra/Material-165.pdf).

La reclasificación de los anaerobios no modificó el rendimiento ni la jerarquía de predictores (ρ 0,95-0,97), mientras que una definición estrictamente microbiológica redujo la discriminación (AUROC 0,699) y desplazó el peso desde el foco relacionado con catéter hacia los marcadores de exposición sanitaria previa (ρ 0,72) (Material suplementario https://revistaemergencias.org/extra/Material-165.pdf).

Discusión

Este estudio desarrolla y valida internamente siete modelos predictivos para estimar la adecuación de la cobertura empírica con ceftriaxona en pacientes con bacteriemia atendidos en el SU, utilizando exclusivamente información disponible durante la valoración inicial. Los siete mostraron una capacidad discriminativa moderada, sin diferencias relevantes entre sí (ABC COR 0,716-0,732), y un beneficio clínico neto similar. La información clínica disponible antes del resultado microbiológico permite estratificar precozmente este riesgo y apoyar la selección del tratamiento empírico.

La bacteriemia es la causa principal de sepsis, uno de los escenarios donde la adecuación del tratamiento inicial más impacta en el pronóstico[4,8]. La precocidad del tratamiento empírico activo ha sido revisada sistemáticamente en España[20]. El modelo no predice la resistencia microbiológica de un aislamiento concreto, sino la pro-

Avance online de artículo en prensa

Tabla 3. Modelo final de regresión logística: coeficientes y *odds ratio* de las variables retenidas

Variable	Coeficiente (β)	*Odds ratio* (IC 95%)
Foco relacionado con catéter	1,289	3,63 (2,33-5,50)
Foco de piel y partes blandas	1,010	2,74 (1,66-4,36)
Infección relacionada con la asistencia sanitaria	0,711	2,04 (1,49-2,83)
Antecedente de colonización previa por MMR (12 meses)	0,640	1,90 (1,35-2,69)
Ingreso en centro sociosanitario	0,441	1,55 (1,01-2,42)
Vive en residencia	0,339	1,40 (1,03-1,96)
Deterioro cognitivo	0,319	1,38 (1,01-1,90)
Sexo masculino	0,218	1,24 (1,00-1,55)
Ingreso hospitalario en el año previo	0,114	1,12 (1,00-1,49)
Uso de quinolonas (3 meses previos)	0,003	1,00 (1,00-1,69)
Edad (por desviación estándar, 15,5 años)	–0,013	0,99 (0,87-1,06)
Foco abdominal	–0,253	0,78 (0,53-1,00)
Foco respiratorio	–1,208	0,30 (0,16-0,48)
Constante	–1,020	–

MMR: microorganismo multirresistente.
La regularización L1 anuló los coeficientes de las 12 variables restantes (diabetes mellitus, EPOC, insuficiencia renal crónica, insuficiencia cardiaca, hepatopatía crónica, neoplasia, trasplante, foco urinario y uso previo de carbapenémicos, β-lactámicos, cotrimoxazol y fosfomicina/nitrofurantoína), que por tanto no intervienen en el cálculo. La edad se introdujo estandarizada; el *odds ratio* corresponde a una desviación estándar. Para el cálculo del riesgo individual la edad se tipifica como (edad – 74,64)/15,60, valores de la cohorte completa; la media del subconjunto de entrenamiento difiere en menos de 0,4 años (IC 95%), con un efecto sobre la probabilidad estimada inferior a 0,0004.

babilidad de que la ceftriaxona no constituya una cobertura empírica adecuada en el momento de iniciar el tratamiento. Por ello se consideró inadecuada la cobertura frente a *Staphylococcus aureus,* incluidas las cepas sensibles a meticilina, ya que el tratamiento con ceftriaxona se relaciona con un aumento de la mortalidad[21].

González del Castillo *et al.*[22] desarrollaron el modelo ATM en 5.170 pacientes de 54 SU españoles para estimar el riesgo de infección por MMR (ABC COR 0,76; sensibilidad 91%), frente al ABC COR de 0,731 y la sensibilidad del 70,6% de RL en este estudio. La diferencia es esperable, ya que ambos estiman desenlaces distintos: estratificación del riesgo global de MMR frente a adecuación de un antibiótico concreto en poblaciones con distintas comorbilidades y diferente prevalencia de MMR.

En cohortes hospitalarias mayores, Vázquez Guillamet *et al.*[23] y Lewin-Epstein *et al.*[24] obtuvieron ABC COR de 0,75-0,76, que aumentaban a 0,80-0,88 al incorporar información microbiológica. El primero atribuyó la variabilidad entre hospitales más a la ecología local que a las características individuales. Corbin *et al.*[25] mostraron que los antibiogramas personalizados mejoran la selección frente a los agregados hospitalarios. En cambio, el presente modelo prescinde de información microbiológica, que no está disponible cuando debe prescribirse el tratamiento empírico en SU. Peyrony *et al.*[26] caracterizaron los factores de riesgo de MMR en pacientes hematológicos atendidos en el SU por sospecha de bacteriemia.

El incremento de la complejidad algorítmica no se tradujo en una mejora clínicamente relevante. Coincide con la revisión sistemática de Christodoulou *et al.*[27], que no halló superioridad consistente del aprendizaje automático frente a la RL con datos clínicos estructurados, y con Gallardo-Pizarro *et al.*[28], que alcanzaron la misma conclusión en el propio SU. Probablemente la información predictiva disponible en la valoración inicial se concentra en pocas variables con elevado poder discriminativo, lo que limita la ganancia de algoritmos capaces de modelizar relaciones complejas. La RL se seleccionó por combinar ese rendimiento con mayor interpretabilidad, estabilidad y facilidad de implementación.

El análisis SHAP mostró elevada concordancia entre ambos modelos. La infección relacionada con la asistencia sanitaria fue el predictor de mayor peso, seguida del foco relacionado con catéter y del antecedente de colonización por MMR, coherentes con los factores de riesgo clásicamente asociados a resistencia[9-11]. El foco respiratorio contribuyó en dirección protectora, en línea con el predominio de microorganismos sensibles en las neumonías comunitarias. La edad tuvo una contribución modesta: la regularización L1 pudo atenuar coeficientes cuya señal capturan variables correlacionadas, como el deterioro cognitivo y la institucionalización, lo que explica también el escaso peso de la exposición antibiótica previa. Debe considerarse además, que en un modelo lineal, los valores SHAP son el producto del coeficiente por la desviación de la variable respecto a su media, de modo que la importancia refleja también su frecuencia.

La interpretación del sexo como predictor, la relación entre la probabilidad umbral del ACD y el umbral operativo, y las ventajas de una puntuación simplificada frente a la ecuación completa, propuesta por Haimerl *et al.*[29], se detallan en el material suplementario (Material suplementario https://revistaemergencias.org/extra/Material-165.pdf).

La divergencia al aplicar una definición estrictamente microbiológica refleja que ambos desenlaces son objetivos de predicción distintos: al excluir *Staphylococcus aureus* sensible a meticilina desaparece la asociación entre el foco por catéter y la inadecuación, mientras que la resistencia en *Enterobacterales* viene determinada por la exposición sanitaria previa.

La contrapartida de priorizar la sensibilidad es un valor predictivo positivo moderado (0,56) y la persistencia de falsos negativos: 40 de los 136 pacientes con cobertura inadecuada no serían identificados. El equilibrio es deliberado, ya que el coste de un tratamiento empírico inactivo supera al de una ampliación reevaluable a las 48-72 horas.

En la cohorte de validación el modelo identificó como de alto riesgo a 172 de 336 pacientes (51,2%). Frente a la ampliación empírica universal, evitaría 164 ampliaciones (49%), a costa de no detectar 40 de los 136 casos inadecuados. Frente al uso indiscriminado de ceftriaxona, evitaría 96 tratamientos inactivos a cambio de 76 ampliaciones innecesarias: 0,79 ampliaciones por cada tratamiento inactivo evitado (0,68 en la validación temporal). Su verificación requiere un estudio prospectivo.

Avance online de artículo en prensa

En la aplicación asistencial, la ausencia de alguna variable no impide el cálculo: los predictores binarios no documentados deben codificarse como ausentes, criterio equivalente al empleado durante el desarrollo, si bien ello puede infraestimar el riesgo en pacientes con antecedentes no registrados.

La prevalencia observada (40,4%) depende de la ecología local, por lo que el intercepto de calibración diferirá en centros con distinto patrón de resistencias, como ilustra la validación internacional de Deelen *et al.*[30]. Los datos de este estudio lo confirman, con un aumento del 38,5% al 45,0% en 3 años en el mismo centro. Todo despliegue externo debería ir precedido de una recalibración local que contemple intercepto y pendiente, con actualización periódica.

En la práctica asistencial, un paciente de 82 años procedente de residencia, con deterioro cognitivo e infección relacionada con la asistencia sanitaria, alcanza una probabilidad estimada del 49%, que asciende al 78% si el foco es un catéter; ambos superan el umbral y apoyarían ampliar la cobertura. Un paciente de 68 años con bacteriemia comunitaria de foco respiratorio se sitúa en el 7%, lo que respalda mantener ceftriaxona.

Son fortalezas de este estudio la comparación homogénea de siete algoritmos bajo un mismo flujo, el uso exclusivo de variables disponibles en la atención inicial y la combinación de SHAP y DCA con estimaciones por remuestreo. Entre las limitaciones, el diseño retrospectivo y unicéntrico y un tamaño muestral moderado pueden haber introducido sesgos y limitan la transportabilidad a hospitales con distinta epidemiología de resistencias. Además, no se realizó validación externa. Se incluyeron conjuntamente bacteriemias por grampositivos y gramnegativos y focos poco frecuentes, sin análisis estratificado, lo que favorece la representatividad clínica del SU a costa de un rendimiento potencialmente variable en subgrupos. El desenlace se definió desde una perspectiva terapéutica: se consideró inadecuada la cobertura frente a *S. aureus* con independencia de su sensibilidad a meticilina, y frente a anaerobios sin contemplar la adición de metronidazol habitual en el foco abdominal, lo que puede sobrestimar la inadecuación en ese subgrupo. Por último, no se realizaron análisis desagregados por sexo, al no permitir estimaciones estables el tamaño de los subgrupos, ni se dispuso de información sobre género.

En conclusión, en pacientes con bacteriemia atendidos en el SU es posible estimar precozmente la probabilidad de que la ceftriaxona proporcione una cobertura empírica adecuada utilizando exclusivamente variables clínicas disponibles durante la valoración inicial. Los siete modelos evaluados no mostraron diferencias relevantes en rendimiento discriminativo ni en utilidad clínica. Cuando distintos algoritmos ofrecen un rendimiento similar, la interpretabilidad, la reproducibilidad y la facilidad de implementación adquieren un papel determinante en su elección. En este contexto, la RL constituye una herramienta transparente y reproducible, y una base razonable para desarrollar un instrumento de apoyo a la decisión clínica en el SU. No obstante, con la capacidad discriminativa observada y en ausencia de validación externa, debe considerarse una propuesta prometedora que requiere validación prospectiva y multicéntrica antes de su aplicación asistencial.

Conflicto de intereses: Los autores declaran que no existen conflictos de intereses en relación con el presente artículo.

Financiación: Los autores declaran la no existencia de financiación con relación al presente artículo.

Responsabilidades éticas: Todos los autores confirman el mantenimiento de la confidencialidad y respeto de los derechos de los pacientes en el documento de responsabilidades del autor, acuerdo de publicación y cesión de derechos a EMERGENCIAS. El estudio fue aprobado por el Comité de Ética de Investigación del centro (Ref. IIBSP-MLA-2025-49).

Inteligencia artificial: Se han empleado herramientas de inteligencia artificial como asistencia en la edición del texto, cuyo contenido ha sido revisado y validado íntegramente por los autores.

Artículo no encargado por el Comité Editorial y con revisión externa por pares.

Bibliografía

1 Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399:629-55.
2 Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect. 2012;18:268-81.
3 Laupland KB, Church DL. Population-Based Epidemiology and Microbiology of Community-Onset Bloodstream Infections. Clin Microbiol Rev. 2014;27:647-64.
4 Paterson DL, Bonomo RA. Extended-Spectrum β-Lactamases: a Clinical Update. Clin Microbiol Rev. 2005;18:657-86.
5 Badia JM, Calbo E, Besoli A, Casas I, Díaz E, Domenech M, et al. Vigilància de les Infeccions relacionades amb l'atenció sanitària de Catalunya (VINCat): informe anual 2024. Barcelona: Servei Català de la Salut; 2025.
6 Van Belkum A, Burnham CAD, Rossen JWA, Mallard F, Rochas O, Dunne WM. Innovative and rapid antimicrobial susceptibility testing systems. Nat Rev Microbiol. 2020;18:299-311.
7 Rhee C, Kadri SS, Dekker JP, Danner RL, Chen HC, Fram D, et al. Prevalence of Antibiotic-Resistant Pathogens in Culture-Proven Sepsis and Outcomes Associated With Inadequate and Broad-Spectrum Empiric Antibiotic Use. JAMA Netw Open. 2020;3:e202899.
8 Prescott HC, Antonelli M, Alhazzani W, Møller MH, Alshamsi F, Azevedo LCP, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2026. Crit Care Med. 2026;54:725-812.
9 Quirante OF, Cerrato SG, Pardos SL. Risk factors for bloodstream infections caused by extended-spectrum β-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*. Braz J Infect Dis. 2011;15:370-6.
10 Goodman KE, Lessler J, Cosgrove SE, Harris AD, Lautenbach E, Han JH, et al. A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum β-Lactamase–Producing Organism. Clin Infect Dis. 2016;63:896-903.
11 Zhao W, Sun P, Li W, Shang L. Machine Learning-Based Prediction Model for Multidrug-Resistant Organisms Infections: Performance Evaluation and Interpretability Analysis. Infect Drug Resist. 2025;18:2255-69.
12 Chiang WC, Chen SY, Chien KL, Wu GHM, Yen AMF, Su CP, et al. Predictive model of antimicrobial-resistant Gram-negative bacteremia at the emergency department. Am J Emerg Med. 2007;25:597-607.
13 Julián-Jiménez A, García DE, García de Guadiana-Romualdo L, Merinos-Sánchez G, Candel González FJ. Modelos predictivos de bacteriemia en el servicio de urgencias: revisión sistemática. Emergencias. 2024;36:48-62.
14 Gamazo del Río JJ, López Izquierdo R, Delgado Cárdenas G, Estella A, García DE, Julián-Jiménez A. Extracción de hemocultivos a pacientes adultos atendidos en el servicio de urgencias con sospecha de infección grave o sepsis. Emergencias. 2025;37:131-40.
15 Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster

Avance online de artículo en prensa

B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.
16 Leclercq R, Cantón R, Brown DFJ, Giske CG, Heisig P, MacGowan AP, et al. EUCAST expert rules in antimicrobial susceptibility testing. Clin Microbiol Infect. 2013;19:141-60.
17 Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. arXiv. 2019.
18 Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. arXiv. 2017.
19 Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Med Decis Making. 2006;26:565-74.
20 Rodríguez MR, Llopis Roca F, Rubio Díaz R, García DE, Julián-Jiménez A. Administración precoz de la antibioterapia empírica en los pacientes atendidos con sospecha de sepsis en el servicio de urgencias: revisión sistemática. Emergencias. 2025;37:44-55.
21 Maraolo AE, Nobile M, Gentile I. Ceftriaxone for methicillin-susceptible Staphylococcus aureus bloodstream infections is associated with increased short-term mortality: a systematic review and meta-analysis. Ann Med. 2026;58:2667672.
22 González Del Castillo J, Julián-Jiménez A, Gamazo-Del Rio JJ, García-Lamberechts EJ, Llopis-Roca F, et al. A multidrug-resistant microorganism infection risk prediction model: development and validation in an emergency medicine population. Eur J Clin Microbiol Infect Dis. 2020;39:309-23.
23 Vazquez Guillamet MC, Liu H, Atkinson A, Fraser VJ, Lu C, Kollef MH. Performance of Risk Models for Antimicrobial Resistance in Adult Patients With Sepsis. JAMA Netw Open. 2024;7:e2443658.
24 Lewin-Epstein O, Baruch S, Hadany L, Stein GY, Obolski U. Predicting Antibiotic Resistance in Hospitalized Patients by Applying Machine Learning to Electronic Medical Records. Clin Infect Dis. 2021;72:e848-55.
25 Corbin CK, Sung L, Chattopadhyay A, Noshad M, Chang A, Deresinksi S, et al. Personalized antibiograms for machine learning driven antibiotic selection. Commun Med. 2022;2:38.
26 Peyrony O, García-Pouton N, Chumbita M, Aiello TF, Monzó-Gallo P, Teijon-Lumbreras C, et al. Epidemiología actual de las infecciones documentadas microbiológicamente y factores de riesgo de multirresistencia en pacientes hematológicos que acuden al servicio de urgencias con sospecha de bacteriemia. Emergencias. 2025;37:274-80.
27 Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.
28 Gallardo-Pizarro A, Peyrony O, Teijón-Lumbreras C, Castiella-Aranzasti A, Martínez-Urrea A, Terrones-Campos C, et al. Evaluación del riesgo de bacteriemia en pacientes con neoplasias hematológicas en el servicio de urgencias: estudio comparativo entre regresión logística y algoritmos de aprendizaje automático. Emergencias. 2026;38:089-97.
29 Haimerl BJ, Encinas R, Justo JA, Kohn J, Bookstaver PB, Winders HR, et al. Optimization of empirical antimicrobial therapy in Enterobacterales bloodstream infection using the extended-spectrum beta-lactamase prediction score. Antibiotics (Basel). 2023;12:1003.
30 Deelen JWT, Rottier WC, Giron Ortega JA, Rodríguez-Baño J, Harbarth S, Tacconelli E, et al. An international prospective cohort study to validate 2 prediction rules for infections caused by third-generation cephalosporin-resistant Enterobacterales. Clin Infect Dis. 2021;73:e4475-83.

Avance online de artículo en prensa