



<https://www.revclinesp.es>

1305 - PREDICCIÓN DE EVENTOS NO SIDA EN PERSONAS QUE VIVEN CON VIH A TRAVÉS DE MACHINE LEARNING

Luis Ramos Ruperto, Ramón Puchades Rincón de Arellano, David Sánchez García, María del Mar Arcos Rueda, Carmen Busca Arenzana, José Ignacio Bernardino de la Serna, Juan González García y José Ramón Arribas López

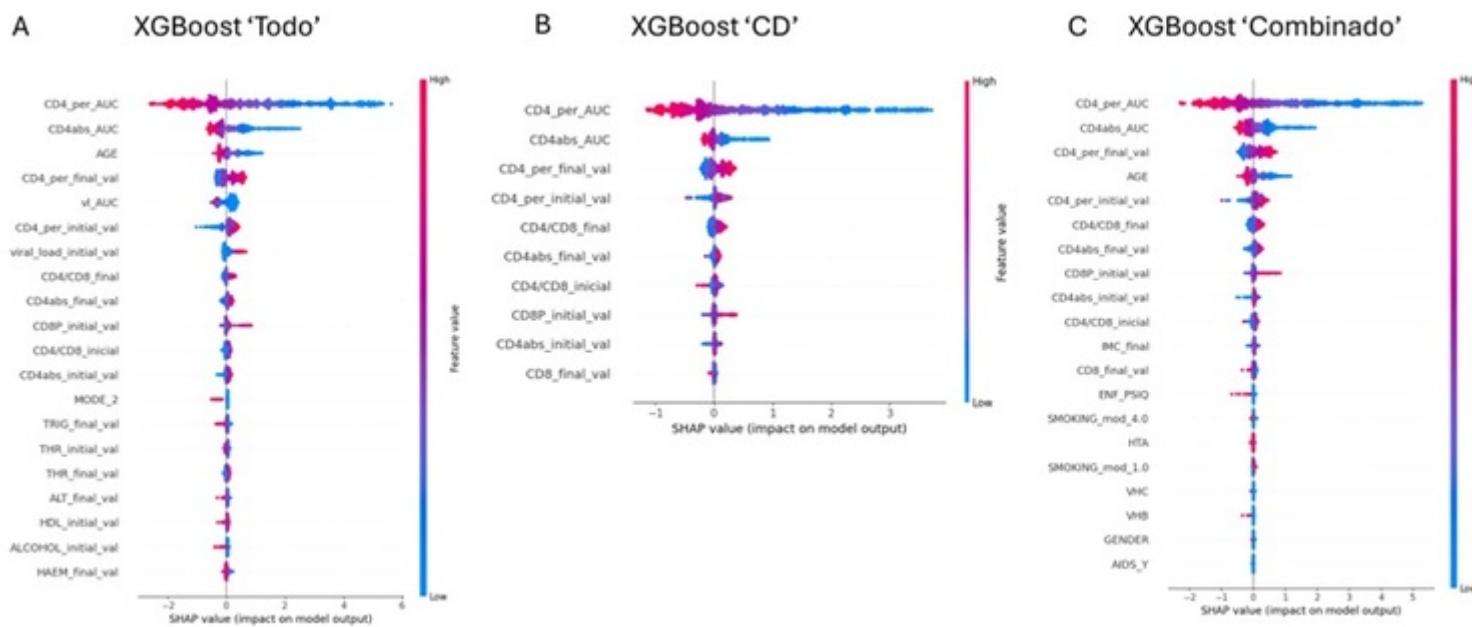
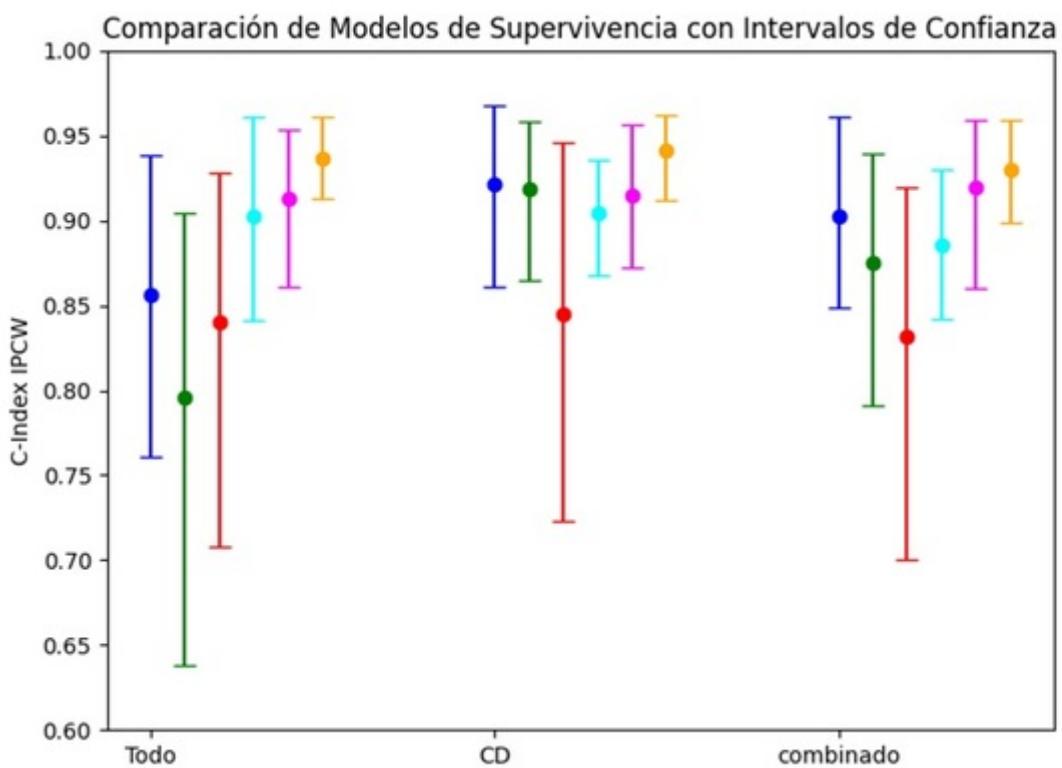
Hospital La Paz-Carlos III-Cantoblanco, Madrid, España.

Resumen

Objetivos: El objetivo principal de este estudio fue desarrollar modelos predictivos utilizando técnicas de *machine learning* (ML) para prever eventos no relacionados con el SIDA en personas que viven con VIH (PVV). Se compararon los modelos ML con las técnicas lineales tradicionales, como la regresión de Cox, para evaluar su eficacia en la predicción de dichos eventos. Además, se aplicaron técnicas de explicabilidad de ML, específicamente los valores SHAP, para comprender mejor las relaciones capturadas por los modelos y su capacidad explicativa.

Métodos: Se llevó a cabo un estudio observacional retrospectivo, revisando las historias clínicas electrónicas de PVV en seguimiento de una Unidad de VIH de un hospital terciario desde 2004 hasta 2022. Se incluyeron pacientes mayores de 18 años, *naïve* al tratamiento antirretroviral al inicio del seguimiento. Se excluyeron aquellos con diagnósticos de VIH a raíz de la ocurrencia del evento combinado y con menos de 180 días de seguimiento. Las variables utilizadas incluyeron datos demográficos, antecedentes médicos, variables relacionadas con la infección por VIH y otras variables adicionales como el índice de masa corporal y el área bajo la curva (AUC) para la carga viral y recuento de CD4. Se preprocesaron los datos eliminando variables con más del 25% de valores nulos y se imputaron datos faltantes utilizando XGBoostRegressor y RandomForestClassifier. Se dividieron los datos en tres particiones: todas las variables (TODO), variables relacionadas con CD4 y CD8 (CD), y una combinación de variables CD y sociodemográficas (Combinado). Para cada partición, se probaron modelos de supervivencia de la biblioteca Scikit-survival, incluyendo regresión de Cox, Random Survival Forest (RSF), Gradient Boosting Survival (GBSA), y XGBoost, adaptado para análisis de supervivencia. La evaluación se realizó mediante el índice de concordancia (c-index) ponderado por la inversa de la probabilidad de censura (IPCW) y se utilizaron técnicas de *bootstrapping* con 100 iteraciones para calcular los intervalos de confianza.

Resultados: XGBoost superó a todos los demás modelos en precisión predictiva en todas las particiones de datos, con diferencias estadísticamente significativas ($p < 0,001$). Especialmente, los modelos restringidos a variables relacionadas con CD4 demostraron un rendimiento superior. El análisis SHAP mostró que la AUC del recuento de CD4 fue una característica dominante, indicando su relación no lineal con el riesgo de eventos no-SIDA. Las características basales de la población mostraron diferencias significativas en variables como el modo de contagio, historia de hepatitis C, hipertensión y recuento final de CD4, entre otros. Los modelos de ensemble como XGBoost demostraron ser mejores para capturar relaciones no lineales entre variables en comparación con modelos lineales tradicionales.



Conclusiones: Los modelos de *machine learning*, especialmente XGBoost, proporcionaron una capacidad predictiva superior para eventos no-SIDA en PVV en comparación con los modelos tradicionales lineales. Las medidas relacionadas con CD4, particularmente la AUC, surgieron como predictores valiosos. Estos hallazgos subrayan el potencial del ML en la mejora de la atención personalizada para PVV, sugiriendo la necesidad de validación en cohortes externas y la expansión para predecir una diversidad de eventos.