



CARTA AL EDITOR

La medicina es frecuentemente desequilibrada



Medicine is often imbalanced

Sr. Editor,

En el artículo de Lozano Gómez et al. «Diseño de un nuevo indicador de mortalidad en el síndrome coronario agudo al ingreso en la Unidad de Cuidados Intensivos»¹ se persigue la ambiciosa tarea de crear un algoritmo de detección de mortalidad en una enfermedad tan relevante como el síndrome coronario. Primeramente, querría reconocer la valía de su trabajo por la inherente complejidad de la base de datos a la que se enfrentan y por la novedosa metodología empleada.

No obstante, me gustaría enfatizar un aspecto que considero no ha sido reflejado adecuadamente en las limitaciones del estudio. La base de datos es muy compleja porque es muy desequilibrada ya que el evento de interés (mortalidad) tiene una prevalencia muy baja (< 5%). Este hecho genera diversos problemas que me propongo exponer.

Las métricas de evaluación de los modelos son importantes porque guían en la elección del mejor modelo posible. Si la guía no es la adecuada, el modelo seleccionado no va a darnos los resultados que buscamos. La métrica utilizada en el trabajo ha sido el área bajo la curva (*receiver operator characteristic* [ROC]), que tiende a ser especialmente optimista en conjuntos de datos desequilibrados². Quizás es por ello que, en el conjunto de datos de validación, encontramos una sensibilidad del 12% y un valor predictivo positivo del 48%. Valores alejados probablemente del objetivo del algoritmo. Sin embargo, muchas veces es difícil utilizar otra métrica (aunque sea subóptima) para no perder capacidad comparativa con otros estudios.

El problema del desequilibrio y de la métrica puede haber sido aumentado por el propio algoritmo usado (Multilayer perceptron)³. Los algoritmos tienden a optimizar el resultado global y muchas veces su optimización interna se encuentra influida por la clase dominante (en el problema actual «vivo»). Técnicamente, el vector de la clase dominante de la variable de interés (vivo) es más «poderoso» durante el *descenso de gradiente* que se obtiene durante la optimización mediante *back propagation*, de tal manera que en bases de datos desequilibradas el algoritmo tiende a optimizar la clasificación adecuada de la clase dominante en el evento de interés (vivo).

Para solventar estos problemas, se proponen métodos que actúan sobre la base de datos y los algoritmos:

- Respecto al conjunto de datos: se proponen técnicas de muestreo que permitan un mayor equilibrio de la muestra, ya sea aumentando el número de casos de interés (fallecimiento), disminuyendo el número de casos de no interés (vivo) o ambas a la vez.
- Respecto al algoritmo: utilizar algoritmos basados en *boosting*, adoptar algoritmos con función de coste y utilizar métodos umbral (que ha sido utilizado en el artículo, pero parece que sin conseguir el efecto deseado).

En resumen, el desequilibrio extremo genera complejos problemas estadísticos que son difíciles de solucionar, la (muy usada) curva ROC puede darnos resultados «engañosos» y en este contexto se requiere una mayor reflexión sobre el preprocesamiento y sobre el algoritmo a usar.

Financiación

Ninguna.

Conflicto de intereses

Los autores declaran que ninguno presenta conflicto de intereses.

Bibliografía

1. Lozano Gómez H, Rodríguez García A, Rodríguez Esteban MÁ, López Ferraz C, Murcia Hernández MDP, Fernández Zapata A, et al. Design of a new mortality indicator in acute coronary syndrome on admission to the Intensive Care Unit. *Med Intensiva*. S2173-5727(23)00028-0. Advance online publication. <https://doi.org/10.1016/j.medine.2023.03.008>.
2. Branco P, Torgo L, Ribeiro RP. A survey of predictive modelling under imbalanced distributions. arXiv. Publicado 7 May 2015 [consultado 6 Feb 2023]. Disponible en: <https://arxiv.org/abs/1505.01658>.
3. Sun Y, Wong AKC, Kamel MS. Classification of imbalanced data: A review. *Int J Pattern Recognit Artif Intell*. 2009;23:687–719.

M. Valiente Fernández

Servicio de Medicina Intensiva, Hospital Universitario 12 de Octubre, Madrid, España

Correo electrónico: mvalientefernandez@gmail.com

<https://doi.org/10.1016/j.medin.2023.05.005>
0210-5691/ © 2023 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.