

Modelo de Redes Bayesianas para la Clasificación de Mortalidad por Ictus tras Primer Ingreso: Un Enfoque en Datos Desbalanceados

Adrián Aparicio Mota

José María López Martín, Juan Manuel García Torrecillas

Fundación de Investigación Biosanitaria Andalucía Oriental (FIBAO). Unidad de Investigación Biomédica, Hospital Universitario Torrecárdenas. Departamento de Matemáticas, Universidad de Almería.



Introducción

En la investigación médica sobre ictus isquémico, se busca entender las relaciones entre variables clínicas y sociodemográficas. Se propone un modelo basado en redes bayesianas, en contraste con los modelos frecuentistas comunes. Estas redes no solo identifican asociaciones entre variables, sino que también manejan la incertidumbre en las mediciones clínicas, ofreciendo una perspectiva más completa y reveladora de la complejidad de los datos clínicos. Este enfoque innovador podría proporcionar perspectivas más profundas y precisas en la comprensión de los patrones en pacientes con ictus isquémico.

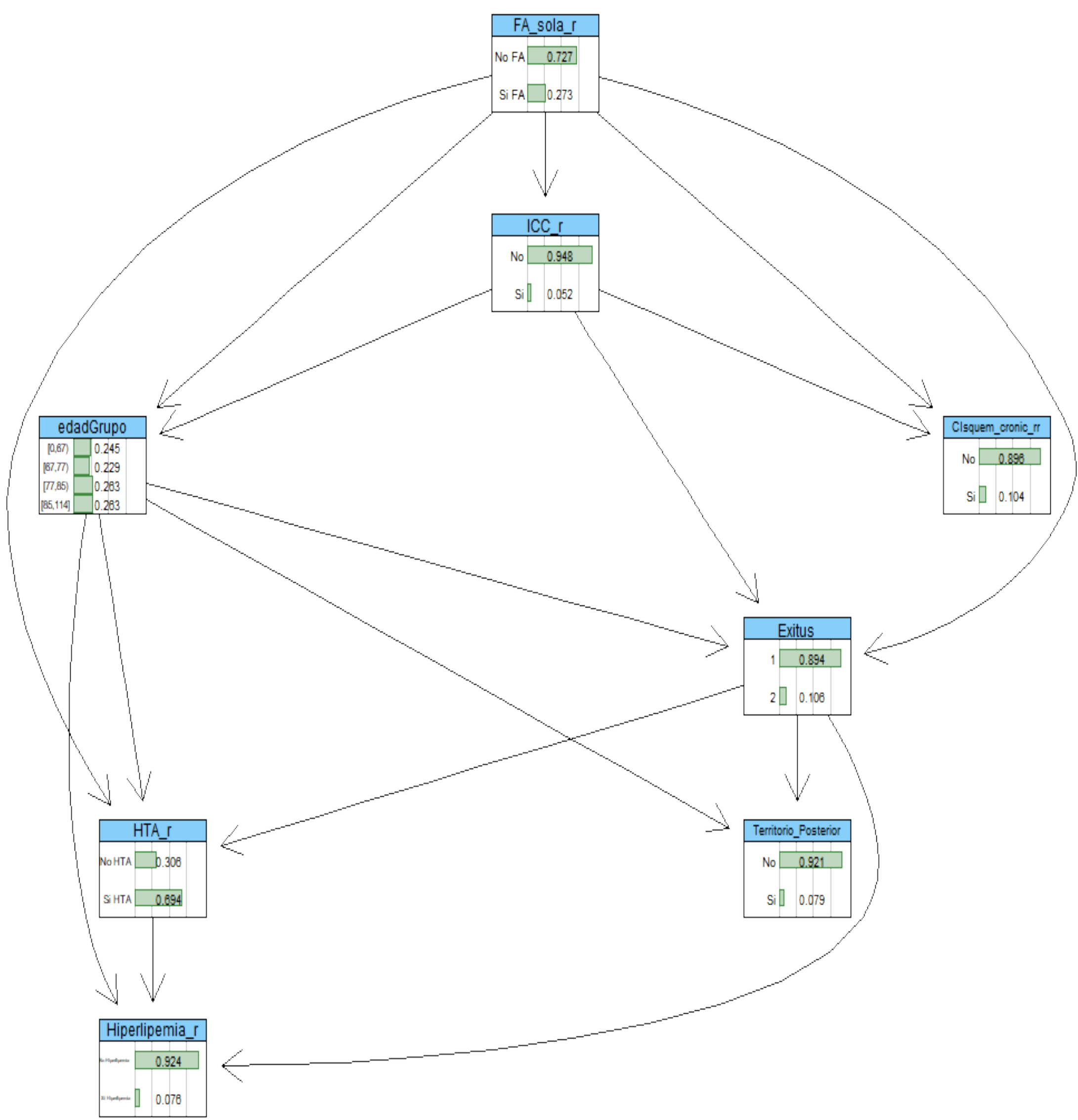
Objetivos

Construir y evaluar qué modelo de red bayesiana tiene la capacidad de representar una estructura para analizar y clasificar la mortalidad (Exitus) en pacientes que ingresan por ictus isquémico en España teniendo conocimientos previos de variables predictoras seleccionadas según bibliografía.

Metodología

Se trabajó con datos de pacientes mayores de 18 años que ingresaron por ictus isquémico en España a partir del Conjunto Mínimo Básico de Datos (CMBD). Se aplicó un modelo de red bayesiana creada mediante algoritmos de aprendizaje de estructura Hill-Climbing mediante puntuaciones aplicando técnicas para el tratamiento de datos desbalanceados.

Estructura de la Red Bayesiana



Tratamiento de datos desbalanceados

1. Oversampling: Consiste en aumentar la cantidad de instancias de la clase minoritaria.
2. Undersampling: Implica reducir la cantidad de instancias de la clase mayoritaria.
3. SMOTE (Synthetic Minority Over-sampling Technique): Genera muestras sintéticas de la clase minoritaria mediante la interpolación entre instancias vecinas.

Fórmula general de la Red Bayesiana

$$P(X) = \prod_{i=1}^n P(X_i | Pa(X_i; \theta_i))$$

Representa la probabilidad conjunta $P(X)$ en una Red Bayesiana, expresada como el producto de las probabilidades condicionales de cada variable X_i dado su conjunto de padres $Pa(X_i; \theta_i)$.

Resultados

Los resultados evidenciaron cómo el manejo de datos desbalanceados puede enriquecer la información en el contexto de un problema de clasificación binaria. En este escenario, se observan métricas más alineadas con el objetivo de clasificar la mortalidad o proporcionar probabilidades de mortalidad más elevadas, evitando así caer en el clasificador trivial que se forma con datos no balanceados (Especificidad=0.998 y Sensibilidad=0.011).

Métricas asociadas al modelo de red bayesiana con tratamiento de datos desbalanceados

	Sensibilidad	Especificidad	VPP	VPN	Accuracy
Sin Balancear	0.011	0.998	0.416	0.894	0.893
Oversampling	0.560	0.776	0.225	0.936	0.748
Undersampling	0.715	0.647	0.194	0.950	0.654
SMOTE	0.821	0.484	0.159	0.956	0.520

Conclusiones

En resumen, la integración de redes bayesianas con estrategias de balanceo de datos se revela como una herramienta crucial para perfeccionar la identificación de casos críticos, especialmente en la clasificación de la mortalidad tras el primer ingreso por ictus isquémico. Esta combinación no solo eleva la precisión del modelo, sino que también fortalece su capacidad para abordar el desequilibrio en los datos, proporcionando así una base más sólida para la toma de decisiones médicas en este contexto específico.

Referencias

¹ PEARL JUDEA. (2000). "Causality: Models, Reasoning and Inference", CUP.

² Nagarajan R, Scutari M (2013). Bayesian Networks in R with Applications in Systems Biology. Springer, New York. doi:10.1007/9781-4614-6446-4, ISBN 978-1-4614-6445-7, 978-1-4614-6446-4.

³ Scutari M, Denis J (2021). Bayesian Networks with Examples in R, 2nd edition. Chapman and Hall, Boca Raton. ISBN 9780367366513.