VALIDACION EXTERNA MEDIANTE TÉCNICAS DE DATA SCIENCE APLICADAS A UN MODELO PREDICTIVO DE MORTALIDAD EN ICTUS ISQUÉMICO

JM García-Torrecillas[1], Mar Iglesias-Espinosa[2], M Mar Rodríguez-Salvador[3], A Aparicio-Mota[4], J Galván-Espinosa[5], MC Lea-Pereira[6]

1.Servicio de Urgencias. Unidad de Investigación Biomédica. Hospital Universitario Torrecárdenas. 2. Servicio de Neurología. Hospital Universitario Torrecárdenas. 3. UGC Ciudad Jardín. Distrito Sanitario Almería. 4. Unidad de Investigación Biomédica. HU Torrecárdenas. 5. Fundación FIBAO. HU Torrecárdenas. 4. Servicio de Medicina Interna. Hospital Universitario de Poniente

CONGRESO INTERNACIONAL DE INVESTIGACIÓN EN SALUD Y ENVEJECIMIENTO & CONGRESO INTERNACIONAL DE INVESTIGACIÓN EN SALUD

Murcia, 20 y 21 de Septiembre de 2022

OBJETIVOS

•El ictus isquémico no reperfundido (INR) corresponde a una clase mayoritaria en cuanto al tratamiento conservador. Un modelo predictivo registrado (MPR, Ref 1) permite conocer su riesgo de mortalidad. Es deseable conocer si este método es útil en pacientes ajenos a aquellos con los que se desarrolló.

•***Objetivo Principal***: Determinar el rendimiento del MPR sobre una cohorte cronológicamente posterior mediante regresión logística (RL). ***Secundarios:*** Evaluar si métodos del data science (logística-balanceada y random-forest) aportan métricas más beneficiosas en la valoración.

METODOLOGÍA

•Fuente de información: Conjunto Mínimo Básico de Datos español, periodos 2008-2012 y 2013-2015 (186.245/114.513 casos), conteniendo todos los ingresados por INR. Diseño y análisis: Estudio de cohortes, análisis de validación mediante aplicación del MPR a base de datos posterior mediante RL; evaluación del modelo mediante RL balanceada, random- forest con/sin undersampling.

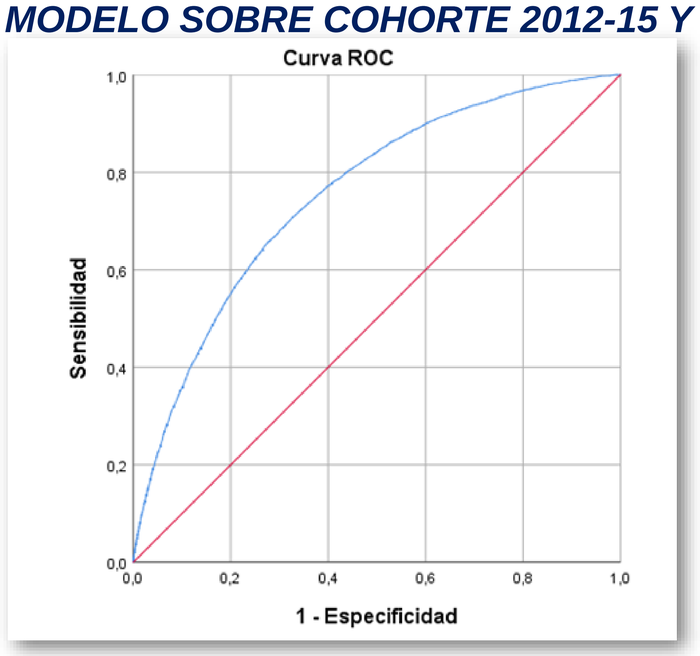
RESULTADOS

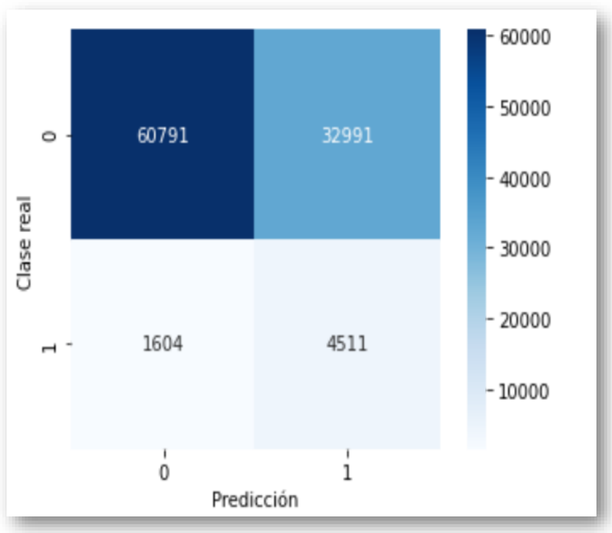
•El modelo original proporcionó un AUC de 0,742, IC95%(0,737-0,747) y mediante RL sobre la cohorte 2013-2015 de 0,754, IC95%(0,749-0,760): diferencia de AUCs <0,10, implicando validación externa eficiente. Métricas: accuracy:0.654, recall:0.738, precision:0.120 y F1-Score:0.207.

•El modelo entrenado sobre la cohorte 2013-2015 mediante data science no mostró ventajas de rendimiento. Los análisis de RL balanceada, random-forest con/sin undersampling y random-forest balanceado, obtuvieron sus mejores resultados en Recall (0.738) siendo deficitarios en precisión (0.120). Los mejores resultados correspondieron al AUC según RL-clásica (0.754-0.756).

GRÁFICAS Y FIGURAS

MODELO SOBRE COHORTE 2012-15 Y TABLA DE CLASIFICACIÓN





MODELO ENTRENADO CON CIENCIA DE DATOS

Comparación de Modelos

```
df_scores
```

	Accuracy	Balanced accuracy	F1-Score	AUROC	Precision	Recall
Balanced Logistic Regression	0.653659	0.689060	0.204882	0.756782	0.119195	0.729397
Random Undersample + Logistic Regression	0.647362	0.689597	0.203846	0.756027	0.118271	0.737722
Balanced random forest	0.654169	0.684565	0.202858	0.754655	0.118097	0.719199

AUC MODELO ORIGINAL: 0,742, IC95% 0,737-0,747

AUC MODELO 2013-2015: 0,754, IC95% 0,749-0,760

DIFERENCIA DE AUCs: 0,013→ VALIDACIÓN EXTERNA CORRECTA Y MÁS PARSIMONIOSA (2 variables menos)

CONCLUSIÓN

Se consiguió la validación externa de un modelo predictivo para mortalidad por INR mediante entrenamiento sobre cohorte posterior. Se evidenció que las metodologías y métricas de procedimientos de data science no aportan beneficio extra en la evaluación.

REFERENCIAS

1. Lea-Pereira MC, Amaya-Pascasio L, Martínez-Sánchez P, et al. JM. **Predictive Model and Mortality Risk Score during Admission for Ischaemic Stroke with Conservative Treatment**. ***Int J Environ Res Public Health***. 2022 Mar 8;19(6):3182

Pertenece al proyecto "Optimización de un modelo predictivo para la mortalidad en ictus isquémico a través de una cohorte de Real World Data". Convocatoria: FPS 2020- Proyectos de I+i en Atención Primaria, Chares y Hospitales Comarcales. Código del Proyecto AP-0013-2020-C1-F1

Junta de Andalucía · Fundación para la Investigación Biosanitaria de Andalucía Oriental - Alejandro Otero · Consejería de Salud y Familias · Fundación Progreso y Salud Consejería de Salud y Familias

Convocatoria FPS 2020. Expdte: AP-0013-2020-C1-F1