



37 - ESTIMACIÓN DE LA PREVALENCIA HOSPITALARIA DE LA DIABETES MEDIANTE APRENDIZAJE AUTOMÁTICO SOBRE TEXTO LIBRE DE HISTORIAS CLÍNICAS ELECTRÓNICAS: ESTUDIO DIABETIC@

A.J. Blanco Carrasco¹, J.F. Navarro², L. Pérez de Isla³, D. Donaire⁴, V. Fanjul⁴, J.J. Aparicio Sánchez⁵, M. Sequera Mutiozaba⁵ y J.F. Merino Torres⁶

¹Servicio de Endocrinología y Nutrición. Hospital Clínic i Provincial de Barcelona. IDIBAPS. Barcelona. ²Servicio de Nefrología. Hospital Universitario Virgen de Candelaria. Tenerife. ³Servicio de Cardiología. Hospital Clínico de Madrid. ⁴Data Science. Savana Research. Madrid. ⁵Departamento Médico Cardiovascular, Renal y Metabolismo. AstraZeneca España. Madrid. ⁶Servicio de Endocrinología y Nutrición. Hospital Universitari i Politècnic La Fe. Valencia.

Resumen

Introducción: La carencia de registros en la mayoría de los sistemas de salud autonómicos limita el conocimiento sobre la población con diabetes *mellitus* (DM) atendida en centros hospitalarios. El estudio Diabétic@, mediante el análisis de datos no estructurados de historias clínicas electrónicas (HCE), pretende generar datos epidemiológicos sobre los pacientes con diabetes tipo 1 (DM1) y tipo 2 (DM2) atendidos en centros de referencia.

Métodos: Estudio observacional basado en el uso secundario de datos no estructurados procedentes de HCE, utilizando tecnología basada en procesamiento del lenguaje natural (PLN) y *machine learning* (ML). Se recogieron datos de 8 hospitales nacionales entre 2013 y 2018. Para inferir el tipo de DM, cuando este no venía reflejado en la HCE, se utilizó un modelo predictivo de ML basado en 107 características clínicas. El rendimiento del modelo se comparó en base a la métrica F.5 (prioriza precisión frente a sensibilidad) con el de un conjunto de 6 reglas clínicas definidas por un comité científico. Finalmente, se estimó la prevalencia de DM1 y DM2 en base a la población clasificada mediante ML.

Resultados: Se analizaron informes de 2.582.778 pacientes, de los que 638.730 tenían DM. El 75,4% de estos no tenía el tipo de DM en su HCE, por lo que fue inferido por el modelo de ML, resultando en 41.028 individuos (6,4%) con DM1 y 588.756 (92,2%) con DM2. El abordaje de ML obtuvo mejores métricas que el de reglas clínicas (F.5 = 0,710 frente a 0,497) y una ratio DM2/1 más alineada con publicaciones recientes (14,3 frente a 3,2). La prevalencia estimada [% (IC95%)] fue 0,5 (0,3-0,7) para DM1 y 14,6 (9-20,2) para DM2.

Conclusiones: El uso de HCE, PLN y ML para identificar y clasificar pacientes con DM se presenta como una herramienta prometedora para obtener el tipo de DM cuando este no viene reflejado en la HCE y con ello obtener una estimación precisa de la prevalencia de DM y estudiar su manejo en entornos hospitalarios.