



6059-4. APLICACIÓN DE HERRAMIENTAS DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DEL DESARROLLO DE CARDIOPATÍA ISQUÉMICA EN PACIENTES CON DOLOR TORÁCICO A PARTIR DE LAS VARIABLES DE LA HISTORIA CLÍNICA ELECTRÓNICA

Alain García Olea¹, Mario Jojoa Acosta², Pablo Osorio Marulanda², Eva Amuriza de Luis³, María Begoña García Zapirain², Mikel Maeztu Rada¹ y Ignacio Díez González¹

¹Hospital Universitario de Basurto, Bilbao, ²Universidad de Deusto, Bilbao y ³Instituto de Investigación Sanitaria Biocruces, Bilbao.

Resumen

Introducción y objetivos: La enfermedad arterial coronaria (EAC) es una de las principales causas de morbilidad. Se ha demostrado que ciertas variables disponibles en la historia clínica electrónica (HCE) pueden introducirse en modelos de regresión para ajustar la probabilidad pretest de desarrollo de EAC en pacientes con dolor torácico. El objetivo del estudio es determinar si la aplicación de herramientas de aprendizaje automático (AA) sobre grandes datos puede mejorar la capacidad predictiva de los modelos convencionales.

Métodos: Utilizamos una muestra de pacientes que requirieron atención médica por dolor torácico o síntomas equivalentes de 2016 a 2018 en centros públicos de la comunidad. Se recogieron 81 variables potencialmente predictoras de EAC y se analizó quiénes desarrollaron EAC en un plazo de un año tras la consulta. La muestra se dividió en un conjunto de 70% de datos para entrenamiento y 30% para validación. Se realizó un proceso de selección de variables con algoritmos de “feature selection”-“random forest” (RF), “support vector machine” (SVM) y regresión logística (RL)-. Se testaron más de 100 algoritmos de AA (redes neuronales, SVM, RL, RF) con diferentes algoritmos de “feature selection” y se ordenaron según su precisión, exactitud, exhaustividad y área bajo la curva (AUC).

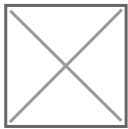
Resultados: 43.835 presentaron dolor torácico en el periodo seleccionado, aunque 10.463 fueron excluidos por no continuarse el estudio. Entre los 33.372 restantes, 5.379 (16,1%) desarrollaron EAC al año. Los cinco algoritmos con mayor potencial predictivo se recogen en la tabla, todos ellos redes neuronales (aprendizaje profundo). Las características de los modelos se recogen en la figura, con un mejor AUC de 0,959 (frente a 0,93 con RL). Las variables con mayor peso fueron la presencia de coronariografía previa, dislipemia, hipertensión, instauración de tratamiento antiisquémico y, de forma inversa, la ansiedad y la fibrilación auricular.

Características de los modelos con mayor AUC

Modelo

AUC

NN-RF-TOP30'	0,95931
NN-Weighted_Intersection25	0,95930
NN-MRMR-TOP30	0,95832
NN-RF-TOP20	0,95676
NN-RF-TOP25	0,95673



Características de los modelos: Modelo, Exactitud, DE, Precisión, DE, Exhaustividad, DE, AUC, DE.

Conclusiones: Los modelos de AA arrojan mejor capacidad predictora que los modelos de regresión convencional en la predicción de desarrollo de EAC a partir de las variables de la HCE. Las principales limitaciones del estudio radican en la mala codificación de los datos estructurados extraídos. Es precisa la validación del algoritmo en una cohorte prospectiva y enfrentarla a los normogramas desarrollados por modelos de regresión para confirmar esta superioridad.