Comentario de literatura destacada

Aplicación de cinco métodos de aprendizaje automático para predecir el nivel de glucosa en sangre a las 52 semanas en pacientes con diabetes tipo 2

Francisco Pérez B[1], Gabriel Cavada Ch[2].*

Implementation of five machine learning methods to predict the 52-week blood glucose level in patients with type 2 diabetes

1. INTA, Universidad de Chile. Santiago, Chile.
2. Escuela de Salud Pública. Facultad de Medicina. Universidad de Chile. Santiago, Chile.

*Correspondencia:
Francisco Pérez / fperez@inta.uchile.cl

Fu X, Wang Y, Cates RS, Li N, Liu J, Ke D, Liu J, Liu H and Yan S (2023) Implementation of five machine learning methods to predict the 52-week blood glucose level in patients with type 2 diabetes.Front. Endocrinol. 13: 1061507. doi: 10.3389/fendo.2022.1061507

Los métodos de aprendizaje automático han desempeñado un papel de apoyo en la generación de modelos a través del aprendizaje y la evaluación de patrones de análisis de datos. Estos modelos ayudan a descubrir correlaciones subyacentes y proyecciones futuras a partir de los datos. En forma habitual, los algoritmos utilizados se diseñan con conocimientos previos y análisis estadísticos. La validez de estos modelos depende de la calidad de los datos recopilados. En esta investigación se realizó un experimento de entrenamiento sobre un conjunto de datos asociados a una cohorte de pacientes y en segundo término se realizó una evaluación en un conjunto de datos de prueba, comparándose cinco diseños de algoritmos.

Este estudio retrospectivo realizado en China, recopiló información multicéntrica de una encuesta de referencia y seguimiento de pacientes con diabetes tipo 2 (2016-2017). Se seleccionaron pacientes de 150 hospitales provinciales incluyendo 1.396 hombres y 1.256 mujeres mayores de 60 años con diabetes tipo 2. El análisis final sólo incluyo 273 fichas completas con datos bioquimicos en las semanas 16 y 52 de seguimiento.

De acuerdo con las directrices de la Sociedad China de Diabetes, se consideró que los pacientes con valores de hemoglobina glicosilada <6,5% en la semana 52 habían logrado la condición de buen control metabólico. A partir de los datos de la encuesta y los datos de seguimiento se incluyó en el modelo indicadores tales: edad, sexo, antecedentes familiares, nivel educacional, evaluación de la dieta, complicaciones micro y macrovasculares (retinopatía, enfermedad renal, neuropatía periférica, aterosclerosis periférica, claudicación intermitente), hipertensión, consumo de alcohol, tabaquismo, IMC, e indicadores bioquímicos tales como: HDL, Hb, K, Na, Cl, CO2, Ca, P, GPT, GOT, rGT, etc). Se construyeron modelos de aprendizaje automático sobre las variables estadísticamente significativas obtenidas del análisis univariante como variables predictoras.

Se modelaron predicciones de glucosa sanguínea utilizando cinco algoritmos de aprendizaje automático (KNN, regresión logística, Random Forest, Support Vector Machine y XGBoost). El análisis de los datos reveló que el algoritmo XGboost obtuvo los mejores resultados que otros modelos, con la mayor precisión en la predicción. Este modelo ajustado podría se de gran ayuda clínica para clasificar

a aquellos pacientes con una alto riesgo de fracaso en el control glicémico, prestar más atención a estos pacientes y orientar su estilo de vida, realizando ajustes diferenciados en su tratamiento y medicación.

Análisis estadístico del estudio

Desde el punto de vista bioestadístico el artículo propone metodologías de predicción bastante vanguardistas basadas en aprendizaje automático. Con resultados bastante alentadores, dado el conjunto de variables que aparecen como predictoras y con ello la plausibilidad biológica de ellas. Sin embargo, llama la atención que el mejor modelo predictivo, tenga una precisión pronóstica del 99.54% en la muestra de entrenamiento y esta baje dramáticamente al 71.18% en la muestra de prueba. Como es sabido, la capacidad de discriminación se cuantifica a través de la Área Bajo la Curva ROC (análisis AUC), que en la muestra de entrenamiento toma el valor 1.0, es decir, una capacidad de discriminación perfecta, y, en la muestra de prueba esta caiga dramáticamente a un 0.68, lo que significa que se gana sólo un 18% al azar (valor de referencia para el azar completo 0.5); según los autores Hosmer y Lemeshow en su libro “Applied Logistic Regression” segunda edición, AUC mayores a 0.7 comienzan a ser valores aceptables para la discriminación, es decir, que el AUC=0.68 reportado por los autores está mas bien cerca de la discriminación pronóstica debida al azar, que en la plausibilidad del modelo. Los autores del artículo concluyen que sus resultados podrían guiar decisiones clínicas posteriores, pero dada la poca capacidad de discriminación, la conclusión es temeraria. Valdría la pena repetir el ensayo con un volumen mas considerable de pacientes, todo indica que 273 pacientes es un número poco decidor o hay que ser mas cuidadoso en los criterios de inclusión y exclusión de los pacientes seleccionados. Por lo tanto, del artículo, parece rescatable la metodología de análisis, pero los resultados no son tan alagüeños.