



## ANÁLISIS CON MODELOS DE APRENDIZAJE AUTOMÁTICO PARA EL DIAGNÓSTICO DE CÁNCER DE PÁNCREAS CON BIOPSIA LÍQUIDA TÉRMICA (TLB)

Sonia Hermoso-Durán<sup>1,2,3</sup>, Nicolas Fraunhoffer<sup>4</sup>, Óscar Sánchez-Gracia<sup>5</sup>, Julia S. Johansen<sup>6</sup>, Astrid Z. Johansen<sup>6</sup>, Sonia Vega-Sánchez<sup>2</sup>, Pablo F. Garrido<sup>2</sup>, Adrián Velázquez-Campoy<sup>1,2,3</sup> y Olga Abian<sup>1,2,3</sup>

<sup>1</sup>Instituto de Investigación Sanitaria de Aragón (IIS Aragón). <sup>2</sup>Instituto BIFI, Universidad de Zaragoza.

<sup>3</sup>CIBEREHD. <sup>4</sup>Centre de Recherche en Cancérologie de Marseille (CRCM), Francia. <sup>5</sup>Universidad de Zaragoza.

<sup>6</sup>Herlev og Gentofte Hospital, Dinamarca.

### Resumen

**Introducción:** El término biopsia líquida térmica (TLB) se refiere al análisis de muestras de suero mediante calorimetría diferencial de barrido (DSC), técnica que proporciona una imagen global de las proteínas séricas y sus interacciones. En estudios previos realizados en nuestro grupo de investigación se ha mostrado el valor potencial de aplicación de la técnica para el diagnóstico y monitorización de pacientes con enfermedades tumorales (cáncer gástrico, pulmón, melanoma y páncreas). Nuestro objetivo en este trabajo fue optimizar el análisis de las curvas del termograma para predecir y diagnosticar sujetos con cáncer de páncreas y discriminarlos de un grupo de control en un conjunto de muestras procedentes el Hospital Herlev og Gentofte de Dinamarca.

**Métodos:** En este estudio colaborativo internacional, se han incluido muestras de suero retrospectivas de población danesa: 321 muestras de pacientes con cáncer de páncreas y 332 muestras de controles sanos. Los termogramas de DSC obtenidos de las muestras informa de la diferencia de capacidad calorífica entre la solución de suero y la solución tampón de referencia en función de la temperatura. El análisis de dichos termogramas (curvas TLB) se ha realizado utilizando un modelo simplificado con temperaturas y un modelo completo añadiendo variables bioquímicas (antígeno CA-19.9, YKL-40, IL-6 y proteína C reactiva). Los modelos de predicción se construyeron aplicando algoritmos de aprendizaje automático basados en regresión, selección de variables con penalización y aplicando validación cruzada, incluyendo técnicas de remuestreo y categorización de variables continuas (pares de temperaturas).

**Resultados:** El modelo completo mostró un área bajo la curva ROC (0,91 frente a 0,75; prueba de DeLong: valor  $p < 0,001$ ) y una especificidad (98% frente a 46%) mayores que el modelo simplificado. En cambio, la sensibilidad fue mayor en el modelo simplificado (86%) que en el modelo completo (77%).

**Conclusiones:** La adición de variables bioquímicas mejora el rendimiento predictivo de los modelos para el análisis de las curvas de TLB. Sin embargo, es necesaria una mayor validación en otras cohortes independientes de pacientes para confirmar su aplicabilidad en entornos clínicos.