

Un modelo de aprendizaje automático basado en la radiómica de la ecografía tiroidea para discriminar entre nódulos benignos y malignos

Francisco Pérez B^{1*}, Gabriel Cavada Ch².

A Machine Learning Model Based on Thyroid US Radiomics to Discriminate Between Benign and Malignant Nodules

A Machine Learning Model Based on Thyroid US Radiomics to Discriminate Between Benign and Malignant Nodules. Guerrisi A y cols. Cancers. 2024; 16: 3775. <https://doi.org/10.3390/cancers16223775>

1. Instituto de Nutrición y Tecnología de los Alimentos (INTA). Universidad de Chile. Santiago, Chile.
2. Escuela de Salud Pública. Universidad de Chile. Santiago, Chile.

Los nódulos tiroideos son un hallazgo muy frecuente en las ecografías tiroideas, en la gran mayoría de los casos se presentan en forma benigna. Sin embargo, existe un porcentaje no menor de nódulos de carácter maligno. Este tipo de análisis de imágenes, además de requerir una mirada experta y precisa, requiere de disminuir al máximo la posibilidad de errores en el diagnóstico. La ecografía y la biopsia con aguja fina son los métodos de diagnóstico más utilizados y fiables disponibles hasta la fecha, pero a veces se presentan ciertas limitaciones al momento de diferenciar los nódulos benignos de los malignos, principalmente en lo que respecta a la ecografía, por la experiencia del operador.

La radiómica, definida como la extracción e interpretación cuantitativa de características a partir de imágenes médicas y el aprendizaje automático que se ha ido generando en el último tiempo, sumado a la disponibilidad más masiva del uso de inteligencia artificial, puede ofrecer una vía bastante promisoría que complementen y mejoren el diagnóstico.

Este trabajo propuso desarrollar un modelo de aprendizaje automático basado en imágenes ecográficas de tiroides para clasificar los nódulos en benignos y malignos. Para cumplir con este propósito se recogieron imágenes de ecografía de 142 sujetos. De estos sujetos, 40 pacientes (28,2%) pertenecían a la clase «maligno» y 102 pacientes (71,8%) pertenecían a la clase «benigno», según el diagnóstico histológico de la aspiración con aguja fina. Este conjunto de imágenes se utilizó para el entrenamiento, la validación cruzada y las pruebas internas de tres modelos diferentes de aprendizaje automático. Se aplicó un enfoque radiómico robusto, bajo la hipótesis de que la característica radiómica podría capturar la heterogeneidad de la enfermedad entre los dos grupos. Se desarrollaron tres modelos consistentes en cuatro conjuntos de clasificadores

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

COMENTARIO DE LITERATURA DESTACADA

de aprendizaje automático (bosques aleatorios, máquinas de vectores de soporte y clasificadores de vecinos más próximos a k) para la tarea de clasificación binaria de interés.

Posteriormente, se puso a prueba el modelo que reportó las mejores aproximaciones en una cohorte de 21 nuevos pacientes.

El mejor modelo (conjunto de bosques aleatorios) mostró un área bajo la curva de características operativas del receptor (ROC-AUC) (%) de 85 (voto mayoritario), 83,7 ** (media) [80,2-87,2], una precisión (%) de 83, 81,2 ** [77,1-85,2], una sensibilidad (%) de 70, 67,5 ** [64,3-70,7], especificidad (%) de 88, 86,5 ** [82-91], valor predictivo positivo (VPP) (%) de 70, 66,5 ** [57,9-75,1] y valor predictivo negativo (VPN) (%) de 88, 87,1 ** [85,5-88,8] (* $p < 0,05$, ** $p < 0,005$) en la cohorte de pruebas internas. Alcanzó una precisión del 90,5%, una sensibilidad del 100%, una especificidad del 86,7%, un VPP del 75% y un VPN del 100% en la cohorte de pruebas externas.

El modelo constituido por cuatro conjuntos de clasificadores de bosque aleatorio logró identificar todos los ganglios malignos y la mayoría consistente de benignos en la cohorte de prueba externa.

Comentario estadístico: Se trata un estudio de discriminación diagnóstica entre nódulos tiroideos malignos y benignos, el

estudio está bien diseñado, ocupa las metodologías clásicas de evaluación diagnóstica y sus respectivas estadísticas de clasificación (sensibilidad, especificidad, valores predictivos positivos y negativos, y, calidad de discriminación mediante el área bajo la curva ROC). La novedad se presenta en que el análisis de las imágenes se hace usando Machine Learning, lo que pone a las técnicas diagnósticas en la línea de esta herramienta que se basa en el aprendizaje automático, que es una rama de la inteligencia artificial y se centra en el desarrollo de sistemas y algoritmos que permiten a las computadoras aprender y tomar decisiones a partir de datos, sin necesidad de ser programadas explícitamente para realizar tareas específicas.

Los resultados encontrados, coinciden con las magnitudes de las estadísticas de buena clasificación conocidas en la literatura biomédica: área bajo la curva ROC del 85%, sensibilidad y especificidad cercanas al 80%, los valores predictivos positivos y negativos, si bien son bastante buenos, pero ellos dependen de las prevalencia clínicas con que se alimentan los modelos. De hecho los autores concluyen en dicho sentido, tal como se puede leer en las conclusiones, que cito.

En términos generales, el artículo es de muy buen nivel metodológico y se recomienda leerlo con detalle.