

Clasificación automatizada de sobreexpresión de HER2 en biopsias digitalizadas de cáncer gástrico, con base de datos chilena

Juan José Alegría^{1,2}, Mauricio Cerda^{2,3}

¹Departamento de Ciencias de la Computación, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

²SCIAN-Lab, Programa de Anatomía y Biología del Desarrollo, Instituto de Ciencias Biomédicas y Centro de Informática Médica y Telemedicina, Facultad de Medicina, Universidad de Chile, Santiago, Chile.

³Biomedical Neuroscience Institute, Santiago, Chile.

Problema

El cáncer gástrico (CG) en Chile tiene una tasa de mortalidad de 20/100.000 habitantes [1]. En 2010 se descubrió que en CG avanzado con sobreexpresión de HER2, un tratamiento con Trastuzumab puede prolongar la sobrevida del paciente [2]. Para determinar la sobreexpresión de HER2, se realiza un examen inmunohistoquímico que es luego analizado por un patólogo [3]. Debido a la subjetividad inherente de este proceso, se hace necesario estudiar maneras para que el análisis de dicho examen sea más objetivo.

Solución

El examen inmunohistoquímico corresponde a una tinción especial aplicada a una biopsia, donde ciertos patrones indican sobreexpresión de la proteína HER2 y tienen una clasificación particular. Dicha biopsia puede ser escaneada y trabajada como una imagen de alta resolución. Así, se propone utilizar técnicas de procesamiento de imágenes para clasificar automáticamente cada región de la biopsia y la biopsia en su total, proveyendo así una herramienta de apoyo al análisis patológico.

Métodos

Las biopsias, provenientes del estudio PRECISO [4], fueron anotadas por 3 patólogos; dichas anotaciones se usaron para entrenar una red neuronal [5] que clasifica la sobreexpresión de HER2. Cada parche es categorizado como No-sobreexpresado (0+/1+), Equívoco (2+) o Sobreexpresado (3+), con lo cual se calcula la clasificación global de la biopsia.

Variables

Las métricas típicas de aprendizaje de máquinas, como exactitud, son sólo aplicables a los parches de entrenamiento. Por ello, otras dos variables cobran mayor relevancia en esta investigación: concordancia inter-patólogos y concordancia de patólogos vs clasificador automatizado.

Hipótesis

La hipótesis que subyace todo este trabajo es que es posible replicar el proceso de diagnóstico realizado por patólogos para sobreexpresión de proteína HER2 en biopsias de cáncer gástrico teñidas inmunohistoquímicamente utilizando algoritmos de procesamiento de imágenes y aprendizaje de máquinas.

Objetivos

Así, el objetivo principal de esta investigación es validar dicha hipótesis y, al mismo tiempo, construir un prototipo de sistema de apoyo al proceso diagnóstico, que permita identificar visualmente qué regiones de la biopsia tienen sobreexpresión de proteína HER2 y la clasificación global de la pieza.

Resultados

Usando la arquitectura de red *Inception-v3*, entrenamos varios modelos de clasificación automática, los cuales obtuvieron buena exactitud ($\sim 90\%$) en los parches extraídos desde las biopsias. Sin embargo, la concordancia inter-patólogos es baja (α de Krippendorff = 0.12), lo cual se reflejó en los resultados entregados por los clasificadores: todas las biopsias fueron etiquetadas como Sobreexpresadas.

Discusión

Siguiendo los consejos de patólogos, se recolectaron nuevas anotaciones para replicar los experimentos. Además, se entrenará un nuevo clasificador que identifique regiones cancerígenas, para sólo allí realizar la clasificación de HER2.

Agradecimientos: Esta investigación fue parcialmente financiada por los proyectos FONDECYT 11161033 y CORFO (16CTTS-66390).

Referencias

- [1] Ministerio de Salud. (2010). *Guía Clínica CÁNCER GÁSTRICO*. Recuperado el 17 de mayo de 2019, desde <https://www.minsal.cl/portal/url/item/722233c6b943cd08e04001011f011d5e.pdf>.
- [2] Bang, Y. J. et al. (2010). Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *The Lancet*, 376(9742), 687-697.
- [3] Bartley, A. N. et al. (2016). HER2 testing and clinical decision making in gastroesophageal adenocarcinoma: guideline from the College of American Pathologists, American Society for Clinical Pathology, and American Society of Clinical Oncology. *American journal of clinical pathology*, 146(6), 647-669.
- [4] Muller, B. et al. (2012). *Observational Study of Perioperative Chemotherapy in Gastric Cancer (PRECISO)*. Identification No. NCT01633203. Recuperado el 17 de mayo de 2019, desde <https://www.clinicaltrials.gov/ct2/show/study/NCT01633203?term=GOCCHI&rank=1>.
- [5] Szegedy, C. et al. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).