



REDES NEURONALES DEL TIPO TRANSFORMER COMO HERRAMIENTA DE CLASIFICACIÓN DE SOBREEXPRESIÓN DE PROTEÍNA HER2 EN BIOPSIAS DE CÁNCER GÁSTRICO.

Diego Muñoz R^{1,2}, Cristian Muñoz B^{1,2,3,4}, Mauricio Cerda V^{2,3,4}, Carlos Navarro C^{1,2,4}

¹Departamento de Ingeniería Eléctrica, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

³Departamento de Ciencias de la Computación, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

²Laboratory of Scientific Image Analysis (SCIAN-Lab), Facultad de Medicina, Universidad de Chile.

⁴Instituto de Neurociencias Biomédicas, Facultad de Medicina, Universidad de Chile.

RESUMEN

El cáncer gástrico ocupa el segundo lugar en términos de causar más muertes en Chile. Esta situación ha impulsado la investigación de varios métodos de detección automática aplicados a muestras de tejido en biopsias. Actualmente, los métodos utilizados se fundamentan en arquitecturas de redes convolucionales (CNN), ampliamente reconocidas por su capacidad de análisis de imágenes. No obstante, en los últimos años han surgido nuevas tecnologías que permiten capturar relaciones contextuales complejas, como los modelos Transformers. Este trabajo busca aplicar y evaluar transformers para mejorar el diagnóstico de cáncer gástrico a través de la detección automatizada de la sobreexpresión de la proteína HER2 en imágenes de biopsias, logrando una mejora de 3% en la exactitud en comparación con CNN.

MATERIALES Y MÉTODOS

En este trabajo, se emplearon 34 biopsias digitalizadas provenientes del estudio "Observational Study of Perioperative Chemotherapy in Gastric Cancer" (PRECISO [1]). Estas imágenes de gran tamaño como se aprecia en la Figura 1, se separaron en entrenamiento y prueba, y se procesaron los datos para adaptarlos a la red neuronal. Se emplearon técnicas para aumentar la cantidad de datos, como rotación y reflexión de imágenes.

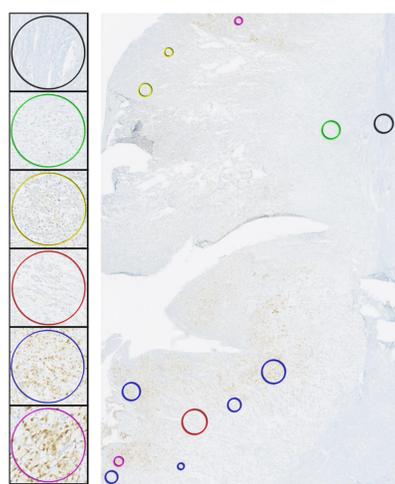


Figura 1: Anotaciones dentro de la imagen de la Biopsia

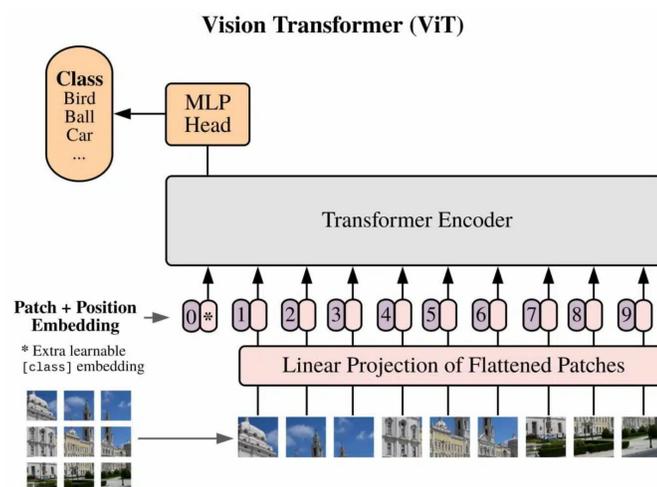


Figura 2: Vision Transformer, obtenido de [3]

El Vision Transformer, como se puede ver en la Figura 2, basa su funcionamiento en descomponer las imágenes en una serie de fragmentos llamados parches, los cuales, luego de ser transformados en vectores, son tratados de la misma forma que las palabras en un Transformer corriente. Todas las capas de atención son del tipo multihead self-attention ya que sólo se utiliza el encoder del Transformer.

Cada anotación experta fue subdividida en parches de 224 x 224, necesarios para la arquitectura transformer. En la figura 3 vemos dos ejemplos en dos escalas distintas, representando diferente cantidad de información.

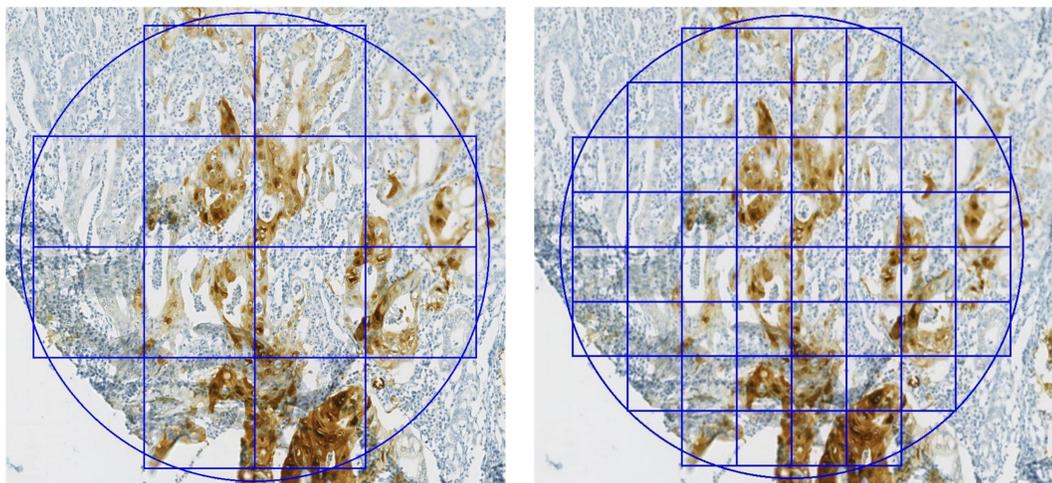


Figura 3: Ejemplos de asignación de parches en dos tamaños a la izquierda, Zoom x10 y a la derecha a Zoom x20

RESULTADOS Y DISCUSION

Los dos métodos basados en Vision Transformer (ViT Small[3] y Swin[4]) superaron en rendimiento a la red convolucional de referencia [2]. El modelo desarrollado logró una mejora de 3% en la exactitud en comparación con la red convolucional, y además, mostró un mejor F1-score en la clasificación de todas las clases. En la tabla 1 se puede ver el primer modelo tumor vs no tumor:

	CNN	ViT Small	Swin
F1 Score No tumor	0.89	0.9	0.92
F1 Score Tumor	0.86	0.88	0.91
Promedio ponderado F1 Score	0.88	0.89	0.91
Exactitud	0.88	0.89	0.91

Tabla 1: F1 Score y Exactitud para Modelo Tumor vs No Tumor

En la tabla 2, es posible apreciar el segundo modelo de nivel de reactividad:

	CNN	ViT Small	Swin
Sin reactividad	0.78	0.79	0.81
React. positiva no lineal	0.36	0.33	0.37
React. casi imperceptible	0.11	0.18	0.25
React. lineal débil	0.55	0.56	0.61
React. lineal fuerte	0.88	0.85	0.88
Promedio ponderado	0.63	0.62	0.65
Exactitud	0.61	0.61	0.64

Tabla 2: F1 Score y Exactitud para Modelo nivel de reactividad.

El trabajo ha demostrado que los modelos ViT son una herramienta prometedora para la clasificación de sobreexpresión de la proteína HER2 en imágenes de biopsias de cáncer gástrico. Es importante mencionar que las CNN pueden presentar ventajas en ciertos escenarios debido a su arquitectura diseñada específicamente para el procesamiento de imágenes. La vinculación interdisciplinaria entre la salud y la ingeniería permite este tipo de soluciones.

REFERENCIAS

- [1] Müller, B., García, C., Sola, J. A., Fernandez, W., Werner, P., Cerda, M., ... & Corvalan, A. H. (2021). Perioperative chemotherapy in locally advanced gastric cancer in Chile: from evidence to daily practice. *ecancermedicalscience*, 15. <https://doi.org/10.3332/ecancer.2021.1244>
- [2] Alegría Fuentes, J. J. (2020). Clasificación automatizada de sobreexpresión de proteína HER2 en biopsias digitalizadas de cáncer gástrico teñidas inmunohistoquímicamente.
- [3] Dosovitskiy, A., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929*.
- [4] Liu, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).

FINANCIAMIENTO Y AGRADECIMIENTOS

Se agradece al Dr. Pablo Zoroquiain por su apoyo anotando las imágenes y discusiones. Este trabajo ha sido parcialmente financiado por FONDECYT 1221696 y FONDEQUIP EQM210020. Esta investigación fue apoyada por el supercomputador Patagón de la Universidad Austral de Chile (FONDEQUIP EQM180042).