

Proyecto: Detección histopatológica del cáncer

Identificación de tejido metastásico en exploraciones histopatológicas de secciones de ganglios linfáticos

✓ Contexto de aplicación

La histopatología es el estudio y diagnóstico de una enfermedad mediante el examen microscópico de una biopsia que se procesa y se fija en un portaobjetos de vidrio. Para visualizar diferentes componentes del tejido, las secciones se tiñen con tintes con el objetivo de revelar los componentes celulares. Los patólogos han utilizado la tinción con hematoxilina-eosina (H&E) durante más de 100 años. La hematoxilina tiñe de azul los tejidos los núcleos celulares, mientras que la eosina tiñe de rosa el citoplasma y el tejido conectivo [1]. Debido a la larga historia del H&E, esta seguirá siendo una práctica común en el análisis de tejidos. En la Figura 1, podemos observar una imagen histopatológica de una sección de ganglio linfático.

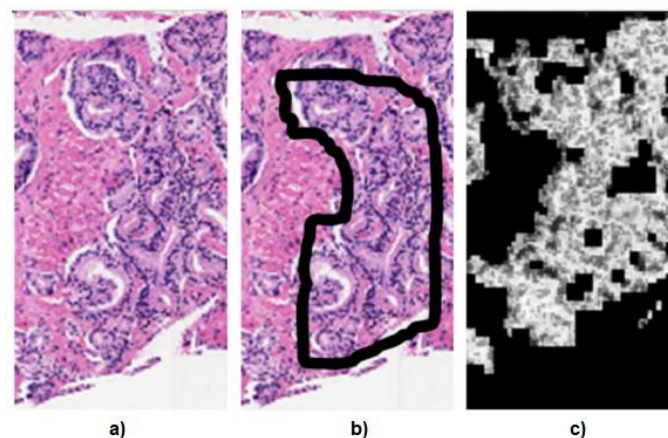


Figura 1. Imagen de histopatología digitalizada, a) original, b) extensión del cáncer anotado por un patólogo en color negro, y c) imagen de probabilidad de cáncer generada por un clasificador *Adaboost*. Imagen tomada de [1].

El sistema linfático es parte del sistema inmunológico del cuerpo, que incluye una red de vasos y ganglios linfáticos. Los vasos linfáticos se parecen mucho a las venas, pero en lugar de transportar sangre transportan un líquido claro y acuoso llamado linfa, el cual contiene glóbulos blancos que ayudan a combatir las infecciones. Los vasos linfáticos envían líquido linfático a través de los ganglios, los cuales son estructuras que actúan como filtros de células cancerosas e infecciones, por lo que combaten, atacan o destruyen dichas infecciones con las células inmunitarias que contienen [2]. El cáncer puede aparecer en los ganglios de dos maneras, puede comenzar allí o puede extender desde otro lugar. Muy frecuentemente el cáncer comienza en otro lugar y luego pasa a los ganglios linfáticos. Esta propagación del

cáncer de un lugar a otro se conoce como metástasis [3], siendo los ganglios más cercanos al cáncer primario los más afectados. Por lo tanto, para determinar si el ganglio linfático tiene cáncer el patólogo examina el tejido teñido procedente de la biopsia y diagnostica la presencia o no de cáncer en los ganglios, basándose en características como la morfología y distribución espacial (Figura 1-b).

Sin embargo, este análisis realizado por el patólogo es laborioso y propenso a subjetividad, por lo que algunos estudios han demostrado un mejor desempeño utilizando modelos de *Deep Learning* en la clasificación histopatológica del cáncer que la realizada por patólogos [4].

✓ **Objetivo de *machine learning***

Para este proyecto se ha seleccionado una competencia de Kaggle sobre detección histopatológica del cáncer¹ la cual estuvo activa desde el 16/10/2018 hasta el 30/03/2019. En este proyecto se debe crear un algoritmo para identificar el cáncer de tejido metastásico en pequeños recortes de imágenes tomadas de exploraciones patológicas digitales más grandes. Los datos son una versión de los datos de referencia de PacthCamelion (PCam)² que a su vez son derivados de los datos de CAMELYON16 [5].

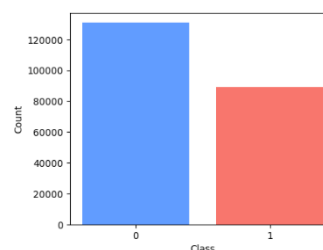
El objetivo del proyecto es predecir la probabilidad de que en la región central (32 × 32 pix) de la imagen, contenga al menos un pixel de tejido tumoral. Por lo tanto, es un problema de clasificación.

✓ **Dataset**

Datos	Tipo	Número de datos	Tamaño en disco
train	.tif	220025	5.87 GB
test	.tif	57458	1.53 GB
train labels	.csv	220025	9.02 MB

▪ **Distribución de las clases:**

Clase	Etiqueta	N° de datos	%
Al menos un pixel con tejido tumoral	1	89117	40.5
Sin tejido tumoral	0	130908	59.5



✓ **Métrica de desempeño**

El modelo de *Deep Learning* creado se evaluará con el AUC (*Area under the receiver operating characteristic curve*) entre la probabilidad predicha y el objeto observado.

¹ <https://www.kaggle.com/competitions/histopathologic-cancer-detection>

² Datos de referencia PCam: <https://github.com/basveeling/pcam>

La curva ROC ilustra la capacidad de diagnóstico del sistema clasificador binario a medida que varía su umbral de discriminación.

✓ **Resultados previos**

Diferentes estudios y competencias han incentivado la creación de algoritmos para la estandarización de imágenes de portaobjetos completas digitalizadas (*whole-slide images*, WSI) y posterior detección de metástasis en los ganglios linfáticos. Una de estas fue la competencia de investigación CAMELYON16 (*Cancer Metastases in Lymph Nodes Challenge 2016*) desarrollada desde nov-2015 hasta nov-2016, donde se proporcionaron datos de entrenamiento y validación de WSI de dos centros de los Países Bajos, que simultáneamente fueron evaluados por 11 patólogos.

A la competencia fueron sometidos 32 algoritmos donde sobresalió por su rendimiento el segundo sometimiento del equipo *Harvard Medical School and Massachusetts institute of Technology* (HMS and MIT II) usando una arquitectura *GoogleNET* (la cual proviene a su vez del reto *ImageNet Large-Scale Visual Recognition Challenge 2014, ILSVRC14*) [6] y creando un nuevo método para la estandarización del color [4], superando tanto el desempeño de los patólogos como del CAD (*Computer-aided diagnosis*). Dentro de las diferentes métricas utilizadas en CAMELYON16 el equipo *HMS and MIT II* obtuvo un AUC de 0.994 y un FROC (*free-response receiver operator characteristic curve*) de 0.807. Por otro lado, otras metodologías han sido creadas para el análisis de WSI en el cual se aprovechan las simetrías de las CNN rotando las imágenes como lo aplican en PCam [5] obteniendo un AUC de 0.963 y un FROC de 0.758.

Respecto al reto asociado a este proyecto, de acuerdo al *leaderboard score*, los ganadores obtuvieron un AUC de 1.0 [7].

✓ **Referencias**

- [1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev Biomed Eng*, vol. 2, pp. 147–171, 2009, doi: 10.1109/RBME.2009.2034865.
- [2] American Cancer Society, "Lymph Nodes and Cancer." Accessed: Sep. 27, 2023. [Online]. Available: <https://www.cancer.org/content/dam/CRC/PDF/Public/7902.00.pdf>
- [3] H. West and J. Jin, "Lymph Nodes and Lymphadenopathy in Cancer," *JAMA Oncol*, vol. 2, no. 7, p. 971, Jul. 2016, doi: 10.1001/jamaoncol.2015.3509.
- [4] B. E. Bejnordi *et al.*, "Stain Specific Standardization of Whole-Slide Histopathological Images," *IEEE Trans Med Imaging*, vol. 35, no. 2, pp. 404–415, Feb. 2016, doi: 10.1109/TMI.2015.2476509.
- [5] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, "Rotation Equivariant CNNs for Digital Pathology," Jun. 2018.
- [6] C. Szegedy *et al.*, "Going Deeper with Convolutions," Sep. 2014.
- [7] W. Cukierski, "Histopathologic Cancer Detection." Kaggle, 2018. [Online]. Available: <https://kaggle.com/competitions/histopathologic-cancer-detection>