

# Clasificación de células en prueba de papanicolaou (pap test) en microscopía\*

Gramática, Martín Nicolás<sup>1</sup>[0009-0009-0269-5778], García, Mario Alejandro<sup>1</sup>[0000-0002-3024-6189], and Gignone, Miguel<sup>1</sup>

Universidad Tecnológica Nacional, Facultad Regional Córdoba, Argentina  
<http://www.frc.utn.edu.ar>

**Abstract.** En este artículo breve se aborda la clasificación de células cervicales en el pap test, también conocido como prueba de Papanicolaou, mediante el uso de YOLOv7. El objetivo final del trabajo es lograr una herramienta que asista a los patólogos en el diagnóstico del cáncer de cuello uterino, mejorando la calidad y los tiempos de diagnóstico. Se realizaron tres experimentos para la obtención de resultados preliminares, utilizando distintos números de clases de entre las seis existentes según el sistema Bethesda (Negativo para lesiones intraepiteliales, ASC-US, ASC-H, LSIL, HSIL, SCC). Se empleó el dataset CRIC para entrenar el modelo, y se desarrollaron una API REST en Flask y una aplicación web para utilizar los modelos entrenados en la inferencia. Los resultados preliminares indican que, con las seis clases, el modelo no logra clasificar con la suficiente exactitud, pero se mejora significativamente al agrupar las cinco clases positivas. Se concluye que las células cervicales son fácilmente detectables con YOLOv7, y que se podrían obtener mejores resultados de clasificación, ya sea mejorando el modelo de detección de objetos, el dataset, o introduciendo un clasificador convolucional como segundo paso.

**Keywords:** pap test · cervical cancer · deep learning · digital pathology

## 1 Introducción

El cáncer de cuello uterino (CCU) es uno de los cánceres más comunes en mujeres a nivel mundial, con una estimación de 604,000 nuevos casos y 342,000 muertes en el año 2020 [1]. En Argentina se diagnostican alrededor de 4500 nuevos casos al año, con una mortalidad de aproximadamente 2000 personas [2].

El diagnóstico precoz del CCU es fundamental para aumentar las posibilidades de supervivencia. Este se lleva a cabo mediante la prueba de Papanicolaou, también conocida como pap test. La realización incorrecta del test, en particular la generación de falsos negativos, puede tener graves consecuencias en la salud de la paciente. Sin embargo, la prueba se ha utilizado ampliamente desde la década de 1950 y ha sido fundamental en la reducción de la incidencia del CCU.

\* Universidad Tecnológica Nacional Facultad Regional Córdoba

En 1988 se estableció un estándar mundialmente aceptado llamado Bethesda (TBS por las siglas en inglés de *The Bethesda System*) [3] para minimizar errores mediante una definición rigurosa de los procedimientos y datos informados en el test. TBS ofrece una categorización general de la muestra, identificando distintos tipos de células, lo cual incluye, de acuerdo a la última actualización que se explica en [4], la clasificación en clases tales como NILM (negativo para lesión intraepitelial), ASC-US (atipia de células escamosas de significado indeterminado), ASC-H (atipia de células escamosas - células escamosas de significado indeterminado), LSIL (lesión escamosa intraepitelial de bajo grado), HSIL (lesión escamosa intraepitelial de alto grado) y SCC (Carcinoma de células escamosas).

Si bien la investigación en métodos automáticos de asistencia al patólogo es un área en constante evolución que ha demostrado ser efectiva en la reducción de errores [5], en Argentina se presentan dificultades para implementar los avances de países desarrollados debido a las diferencias en la tecnología utilizada para la realización de los tests y en la obtención de imágenes. Mientras que los países desarrollados utilizan la citología de base líquida y escáneres WSI para obtener imágenes, en Argentina se sigue utilizando el extendido convencional y microscopios estándares respectivamente.

En este artículo se presenta una idea en desarrollo, junto a resultados preliminares, que consiste en la creación de un sistema sencillo y de fácil disponibilidad para patólogos con el fin de facilitar la elaboración del informe de TBS. Este sistema permitiría la clasificación de células cervicales, lo cual puede mejorar el diagnóstico patológico del CCU.

## 2 Metodología y resultados preliminares

Se realizaron tres experimentos preliminares para la clasificación de células cervicales en pap test. En los tres experimentos se utilizó la arquitectura de red neuronal YOLOv7 [6] para la detección y clasificación de células.

En el primer experimento, se clasificaron las células en seis clases. En el segundo experimento, en dos: células normales (NILM) y alteradas (ASC-US, ASC-H, LSIL, HSIL y SCC). En el tercer experimento, se utilizó solo una clase para detectar la presencia de células cervicales con el único objetivo de identificar células de interés para el procesamiento.

Se utilizó el dataset CRIC [7], el cual contiene 400 imágenes de pap con una resolución de  $1376 \times 1200$  y un archivo json que detalla la ubicación y clase de los núcleos celulares en cada imagen.

Para entrenar el modelo YOLOv7 se crearon archivos de etiquetas asociados a cada imagen. En estos archivos se definieron los bounding boxes utilizando un tamaño fijo de  $90 \times 90$ , tal como en [8]. Las imágenes se recortaron a un tamaño de  $640 \times 640$  píxeles (resolución utilizada por YOLOv7) y por cada imagen se obtuvieron 6 subimágenes con cierto grado de solapamiento para evitar pérdida de información en los bordes. Esto es particularmente importante en este caso debido al tamaño (pequeño) de los objetos a detectar.

En la figura 1 se muestra un resumen del preprocesamiento efectuado sobre el dataset.

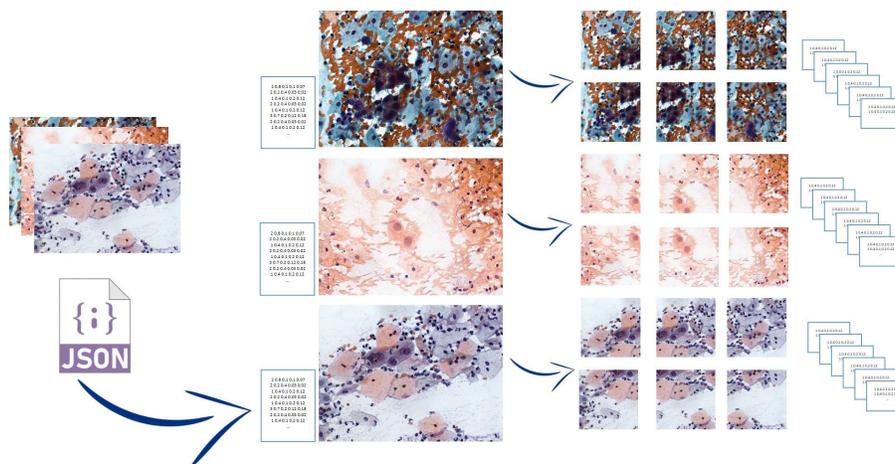


Fig.1: Preprocesamiento del dataset: Partiendo del archivo json que contiene las etiquetas de todas las imágenes se creó un archivo de etiquetas por imagen. Luego, se recortaron 6 imágenes de  $640 \times 640$  por cada imagen y se generaron los archivos de etiquetas correspondientes para cada subimagen.

El conjunto de datos se dividió en particiones de entrenamiento, validación y prueba en una proporción de 80-10-10 sobre las imágenes recortadas, pero teniendo en cuenta de que todas las subimágenes correspondientes a la misma imagen se encuentren en la misma partición para no afectar la evaluación del modelo. La distribución de imágenes es entonces de 1920 para train, 240 para test y 240 para validación. En la tabla 1 se muestra un resumen de la cantidad de objetos por partición para cada uno de los experimentos.

Aunque se observa un evidente desbalance de datos en el experimento con 6 clases, resulta complejo abordar este problema mediante el aumento artificial de datos debido a la notable disparidad en la cantidad de ejemplos disponibles. En concreto, la clase con mayor cantidad de objetos (Negativo) cuenta con 36 veces más ejemplos en el conjunto de entrenamiento en comparación con la clase con menor cantidad (SCC). No obstante, no descartamos la posibilidad de que esta técnica pueda resultar útil, por lo que planteamos analizar su efectividad en futuros experimentos (consultar la sección 4).

Durante el entrenamiento se utilizó un *batch size* de tamaño 4 por limitaciones de Hardware. El modelo se entrenó durante 300 épocas en una GPU NVIDIA GeForce RTX 2080, aplicando fine-tuning sobre un modelo pre-entrenado con el dataset MS COCO.



El modelo resultante fue desplegado en una API REST usando Flask (Python). También se creó una aplicación web sencilla que permite la detección y clasificación en imágenes de  $640 \times 640$  píxeles en tiempo real. De esta manera, el modelo se podrá utilizar de manera práctica y eficiente para la clasificación de células cervicales en la nube.

En la figura 2 se muestra un ejemplo de clasificación utilizando la aplicación web desarrollada.

El modelo logró un  $mAP_{0.5}$  (*mean Average Precision para Intersection over Union = 0.5*) del 60% en la clasificación de células en pap test para el caso de las seis clases y un  $mAP_{0.5}$  del 83% para dos clases. Estos resultados pueden deberse a muchos factores, como por ejemplo que las clases positivas sean muy similares entre ellas, o que el dataset pueda estar desbalanceado. Usando solamente una clase para la detección se logró un  $mAP_{0.5}$  del 84%, lo cual sugiere que YOLOv7 detecta muy satisfactoriamente las células cervicales, pero su rendimiento decrece cuando la tarea es clasificarlas en las 6 clases planteadas.

### 3 Conclusión

Demostramos que YOLOv7 es eficaz en la identificación de células cervicales en el test de Papanicolaou con el conjunto de datos CRIC, pero no así en su clasificación. Es necesario seguir trabajando para mejorar la clasificación de las células detectadas.

Este estudio preliminar representa un importante paso hacia la automatización de la clasificación de células cervicales en pap-test en Argentina, y proporciona una base sólida para futuras investigaciones y mejoras en este campo de estudio.

### 4 Trabajo futuro

Algunos de los trabajos de mejora considerados:

1. Realizar un estudio más detallado del conjunto de datos CRIC para identificar posibles sesgos y desequilibrios en las clases y explorar técnicas de aumento de datos.
2. Realizar un aumento de datos adecuado para proporcionar más variedad de imágenes para el entrenamiento.
3. Investigar la posibilidad de combinar el clasificador YOLO con otro clasificador en un segundo paso, como un modelo basado en redes neuronales convolucionales.
4. Mejorar la generación del dataset, reemplazando los bounding boxes de tamaño fijo, por bounding boxes adaptados a cada célula.
5. Mejorar la aplicación web para que funcione con imágenes de cualquier tamaño, haciendo automáticamente el ajuste de tamaño.
6. Incorporar la capacidad de ejecutar la clasificación en tiempo real, tomando como entrada un stream de video en lugar de imágenes estáticas.

7. Someter los resultados a evaluación por parte de profesionales médicos (Patólogos).
8. Investigar la posibilidad de aplicar técnicas de explicabilidad de modelos, como Grad-CAM o LIME.

## References

1. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3), 209-249.
2. Programa Nacional de Prevención de Cáncer Cervicouterino. Datos epidemiológicos [Internet]. Buenos Aires: Ministerio de Salud de la Nación; [citado 24 feb 2023]. Disponible en: <http://www.msal.gov.ar/inc/el-inc-y-sus-programas/programa-de-prevencion-del-cancer-cervicouterino/#Pautas-para-equipos-de-salud>.
3. The 1988 Bethesda system for reporting cervical/vaginal cytologic diagnoses: Developed and approved at the National Cancer Institute workshop in Bethesda, Maryland, December 12–13, 1988, *Human Pathology*, Volume 21, Issue 7, 1990, Pages 704-708, ISSN 0046-8177, [https://doi.org/10.1016/0046-8177\(90\)90029-5](https://doi.org/10.1016/0046-8177(90)90029-5)
4. Dra. María Carolina Moreno Barrios: Actualización en el reporte de citología cervicovaginal basado en el Sistema Bethesda 2014. *Rev Obstet Ginecol Venez* 2017;77(1): 58 - 66. Venezuela (2017)
5. Zhu, X., Li, X., Ong, K., Zhang, W., Li, W., Li, L., ... & Ding, Y. (2021). Hybrid AI-assistive diagnostic model permits rapid TBS classification of cervical liquid-based thin-layer cell smears. *Nature communications*, 12(1), 3541.
6. Wang, Chien-Yao and Bochkovskiy, Alexey and Liao, Hong-Yuan Mark (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 2022
7. Rezende, M.T., Silva, R., Bernardo, F.d.O. et al. Cric searchable image database as a public platform for conventional pap smear cytology data. *Sci Data* 8, 151 (2021). <https://doi.org/10.1038/s41597-021-00933-8>
8. N. Diniz D, T. Rezende M, G. C. Bianchi A, M. Carneiro C, J. S. Luz E, J. P. Moreira G, M. Ushizima D, N. S. de Medeiros F, J. F. Souza M. A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification. *Journal of Imaging*. 2021; 7(7):111. <https://doi.org/10.3390/jimaging7070111>