



Clasificación de células sanguíneas a partir de imágenes, identificación de células y enfermedades sanguíneas

Procesamiento de imágenes

Darwin Eduardo Martinez Riaño

Esteban Rodríguez - esteban.rodriguez02@usa.edu.co  
Daniel Velasquéz - daniel.velasquez02@usa.edu.co  
Santiago Niño - santiago.nino01@usa.edu.co

Noviembre 17, 2023

## **Resumen**

En la realización de este proyecto se plantearon pasos para el tratamiento de imágenes de células sanguíneas, donde se obtuvieron imágenes de un dataset público, se eligieron cinco (5) imágenes específicas donde se evidencian plaquetas, glóbulos blancos y glóbulos rojos, esto con fines de procesamiento posterior. Ahora bien, una vez seleccionadas las imágenes de prueba, se establecen procesos de procesamiento de imágenes, tal como, reconocimiento de canales de color, transformada de Fourier, filtros de mediana para eliminación de ruido, filtro bilateral, análisis de histogramas de frecuencia, segmentación de histograma, umbralización de imágenes, transformada de Watershed, entre otros procesos, todos estos son procedimientos establecidos para la clasificación de células, determinando, reconociendo y diferenciando los globulos blancos de los rojos, y estos de las plaquetas.

Como enfoque a proceder con el paso anterior logrado, se tuvo previsto implementar las técnicas aplicadas en adición a nuevas técnicas para la detección de enfermedades que se pueden detectar por medio de la sangre, tal como la anemia ferropénica. Para esto se obtuvieron una serie de imágenes de células sanguíneas con el fin de corroborar la eficacia de la detección. Esto se logra mediante la detección de las células, su conteo y análisis de morfología. Este fue el enfoque aplicado a nuestro proyecto, determinar mediante procesamiento de imágenes, la capacidad de detección de enfermedades sanguíneas en base a las células.

## **Palabras Clave**

Segmentación de Imágenes, Umbralización, Extracción de Características, Conteo de células, Clasificación celular, Detección de Enfermedades Sanguíneas.

## **Introducción**

La intersección entre la medicina y la tecnología ha revolucionado la forma en que se abordan los diagnósticos y tratamientos médicos en la actualidad. Uno de los campos de mayor avance en esta convergencia es el análisis automatizado de imágenes, el cual desempeña un papel esencial en la detección temprana y el diagnóstico preciso de diversas enfermedades (Kumar et al., 2019; Hoffbrand et al., 2016). Entre las aplicaciones más destacadas de esta tecnología, la clasificación automatizada de células sanguíneas a partir de imágenes microscópicas ha emergido como una herramienta crucial en el diagnóstico de enfermedades hematológicas.

El diagnóstico tradicional de trastornos sanguíneos ha dependido durante mucho tiempo de la observación manual de extensiones sanguíneas bajo un microscopio (Smith et al., 2018). Aunque este enfoque ha sido fundamental en la detección y el diagnóstico de enfermedades sanguíneas, presenta desafíos significativos. En primer lugar, la observación manual está sujeta a errores humanos, ya que la precisión del diagnóstico puede verse afectada por la fatiga, la experiencia del observador y la variabilidad del interobservador. Además, este proceso puede ser lento

y requerir una inversión considerable de tiempo, lo que puede retrasar el inicio del tratamiento adecuado para los pacientes que sufren de enfermedades sanguíneas. Dada la necesidad de mejorar la eficiencia y la precisión en el diagnóstico de trastornos sanguíneos, se ha producido un crecimiento significativo en la investigación y desarrollo de tecnologías de automatización en este campo. Estos avances han llevado al desarrollo de sistemas de análisis de imágenes microscópicas que utilizan algoritmos y técnicas de procesamiento de imágenes para clasificar automáticamente las células sanguíneas en categorías específicas (Gomez et al., 2020).

El desafío clave en esta área reside en desarrollar sistemas robustos que puedan reconocer y diferenciar con precisión las características de las células sanguíneas en imágenes microscópicas, incluso en presencia de variaciones y condiciones diversas. Esto implica la identificación de glóbulos rojos, glóbulos blancos y plaquetas, así como la detección de cualquier anomalía morfológica que pueda ser indicativa de una enfermedad específica (Johnson et al., 2021).

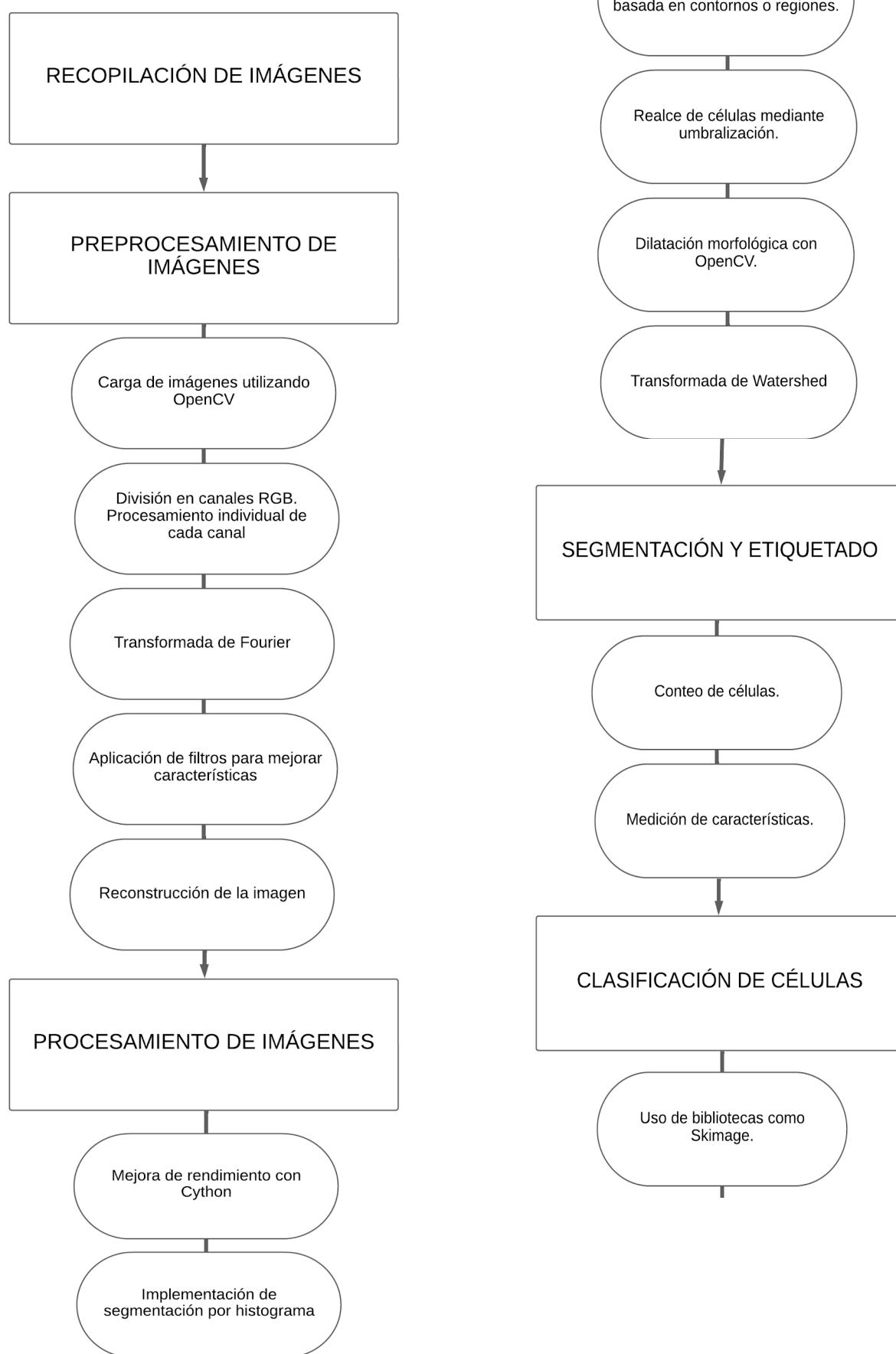
## **Metodología**

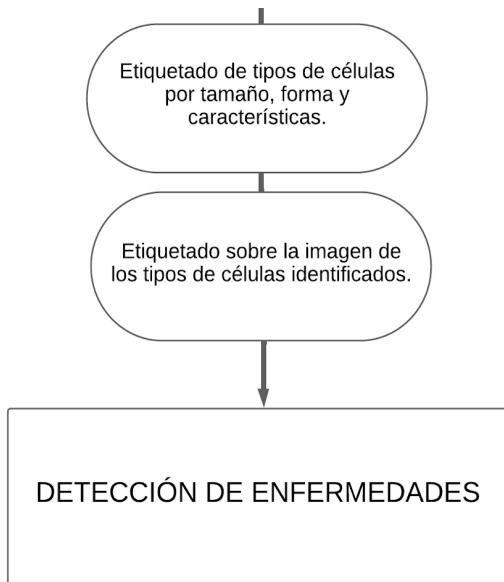
### **Descripción de los datos**

Para el desarrollo del procesamiento de imágenes, se empleó en una primera instancia un dataset público de imágenes de células sanguíneas. Estas imágenes cuentan con un formato JPEG (Joint Photographic Experts Group), debido a su compresión eficiente, las imágenes JPEG tienden a procesarse más rápidamente en comparación con formatos de imagen sin pérdida como PNG o TIFF. Esto puede ser crucial en la aplicación prevista con este proyecto. Ahora bien, estas imágenes cuentan con un tipo de imagen JPEG y un promedio de 640 x 480 píxeles en términos de ancho por alto, respectivamente. Fueron seleccionadas cinco (5) imágenes de este dataset, estas imágenes fueron seleccionadas bajo nuestro criterio destacando el hecho de que las células sanguíneas son más fáciles de diferenciar visualmente, esto se determinó con el fin de mejorar el desempeño de la clasificación activa de las células.

Ahora bien, para la segunda parte de nuestro proyecto, la cual es la detección de enfermedades que pueden ser detectadas mediante la sangre, fue necesario obtener más imágenes de prueba, en este paso se obtuvieron imágenes de pacientes confirmados con estas enfermedades en la sangre, esto con el fin de poner a prueba nuestro procesamiento y comprobar si la detección de enfermedades se realiza de forma correcta y determinar el porcentaje de acierto. Para esta sección, se determinó únicamente a la anemia ferropénica, por fines de alcance del proyecto, como la enfermedad a detectar. Para la anemia, se seleccionaron tres imágenes diferentes, todas son de pacientes confirmados con esta enfermedad, la diferencia radica en el formato de imagen, se manejaron dos imágenes en formato jpg (promedio de 1200 x 1000 píxeles) y una en formato tiff (promedio de 2000 x 1800 píxeles).

## Pipeline de procesamiento





### Descripción del pipeline

La identificación de tipos de células sanguíneas a partir de imágenes se puede realizar utilizando técnicas de procesamiento de imágenes y análisis de imágenes. Estos procedimientos se llevan a cabo en varias etapas clave:

- Recopilación de imágenes:  
En primer lugar, se requirió un conjunto de imágenes de células sanguíneas para el análisis. Estas imágenes se obtuvieron utilizando conjuntos de datos disponibles públicamente que contenían imágenes de células sanguíneas.
- Preprocesamiento de imágenes:  
En esta sección se cargaron las imágenes al proyecto empleando OpenCV, luego se realizó una división de las imágenes en sus canales de color (RGB), se realiza un procesamiento por separado a cada canal, esto con el fin de asegurar el procesamiento equitativo a cada aspecto de las imágenes. Después se aplica la transformada de Fourier, esto se hace para trabajar con las propiedades de frecuencia de la imagen. Se aplican filtros para la mejora de las características de la imagen tratada, por ejemplo, se aplica un filtro de mediana con kernel para eliminar el ruido y se aplica un filtro bilateral para preservar los detalles relevantes. Por último, se reconstruye la imagen, con el ruido reducido y con las características realizadas.
- Procesamiento de imágenes:  
Se empleó Cython para la mejora de rendimiento en el procesamiento de los histogramas, paso siguiente se implementó un algoritmo de segmentación por histograma. La segmentación es un paso crucial en el análisis de imágenes de células sanguíneas, donde se busca separar las células individuales en la imagen. Este proceso se

llevó a cabo mediante diversas técnicas, como la segmentación por umbral, segmentación basada en contornos o segmentación basada en regiones. Luego, se realizaron métodos para el realce de las células en las imágenes mediante umbralización. Paso siguiente, se procede a realizar un proceso de dilatación morfológica de las imágenes umbralizadas, este proceso se realiza mediante el uso de OpenCV, y se realiza para obtener mejores resultados y completos de las células segmentadas. Despues de este proceso, se aplica la transformada de Watershed, este proceso se realiza para la segmentación de regiones de interés de las imágenes, aquí se aplica un desenfoque Gaussiano, un umbral adaptativo y por último, se realizan procesos de erosión y dilatación para realzar detalles, formas y en general, a las células.

El componente de segmentación y etiquetamiento, fue esencial en nuestro proyecto, ya que proporciona una base sólida para el análisis y la identificación precisos de células sanguíneas en imágenes médicas. Las imágenes segmentadas se utilizaron para contar células, medir características y diagnosticar enfermedades relacionadas con la sangre.

- Clasificación de células:  
Se aplica el uso de bibliotecas como Skimage con el propósito de etiquetar regiones específicas de la imagen. Se realiza la identificación de los tipos de células, mediante su tamaño, forma y demás características relevantes. Por último, se realiza un etiquetado sobre la imagen de los tipos de células identificados. Se tienen en cuenta aspectos como el conteo de cada tipo de células que nos va a servir para el desempeño de la identificación de enfermedades.
- Detección de enfermedades:  
Para la detección de la anemia ferropénica.

### Resultados

- Histograma de frecuencia

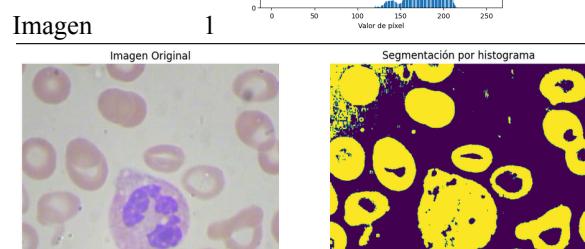
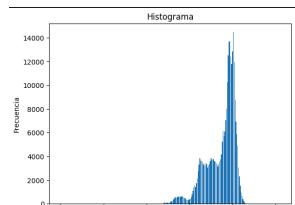


Imagen 2

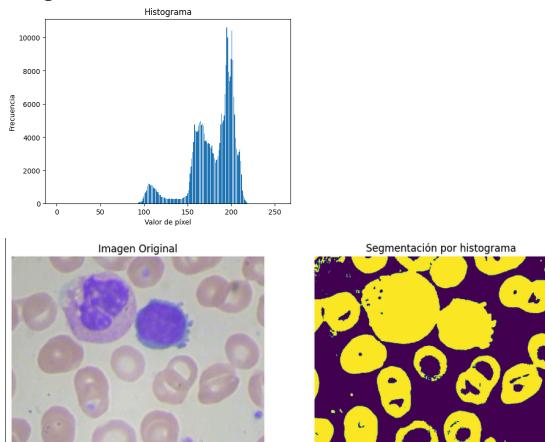


Imagen 3

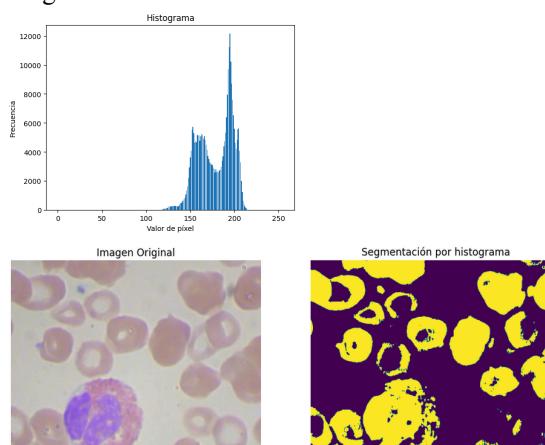


Imagen 4

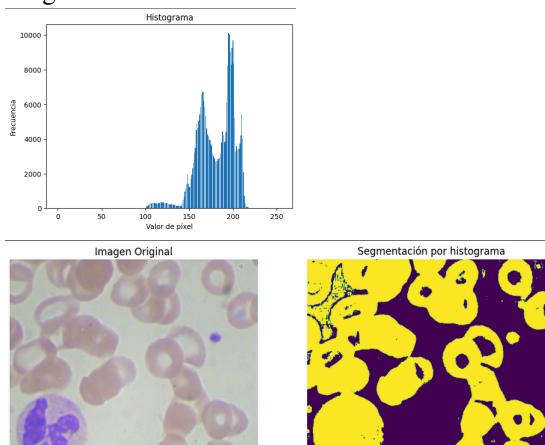
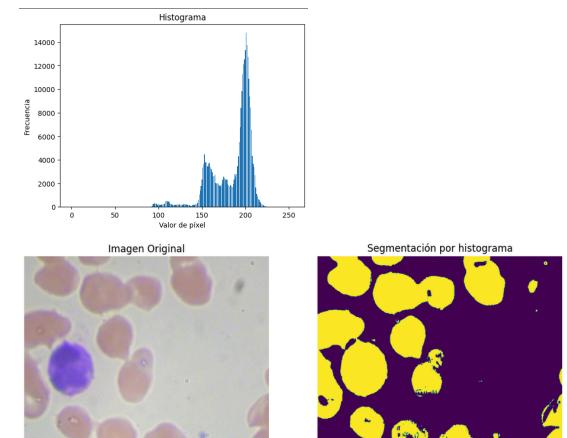


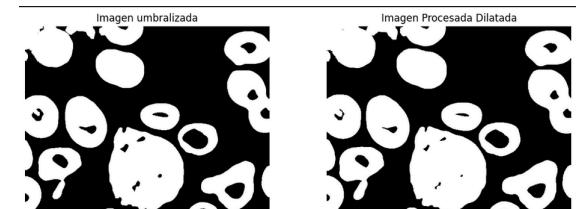
Imagen 5



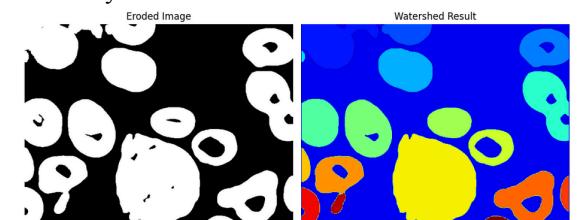
- Procesamiento primera imagen  
Umbralización



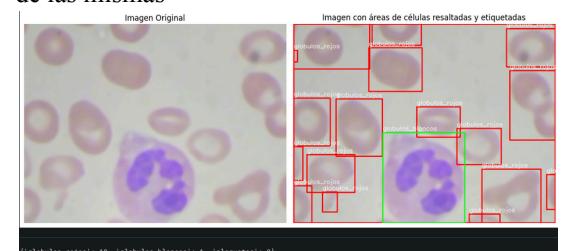
Dilatación de umbralización



Erosión y transformada de Watershed

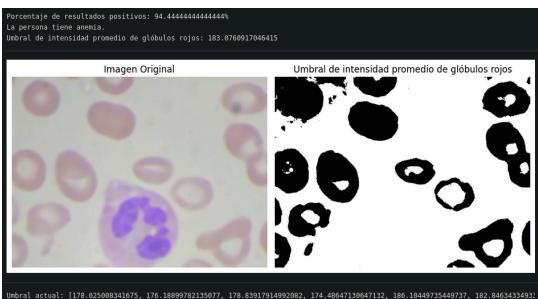


Mostrar etiquetas de tipos de células y conteo de las mismas

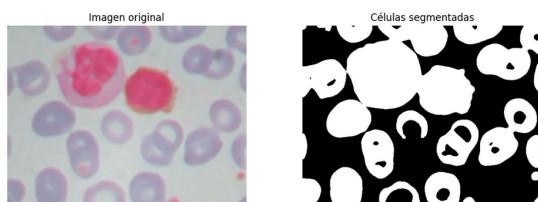


Umbral promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico

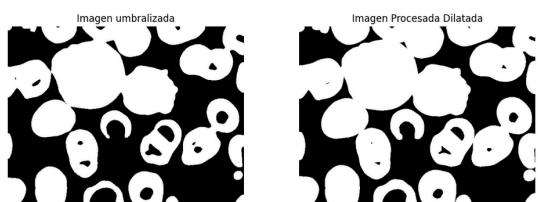
```
['glóbulos_rojos': 10, 'glóbulos_blancos': 1, 'pláquetas': 0]
```



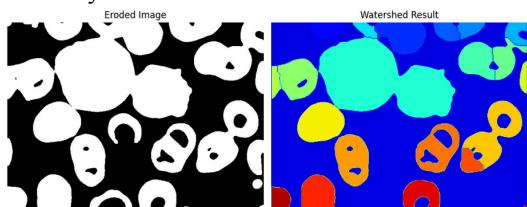
- Procesamiento segunda imagen  
Umbralización



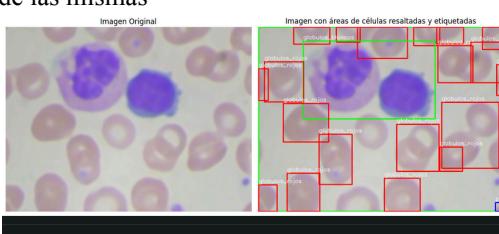
Dilatación de umbralización



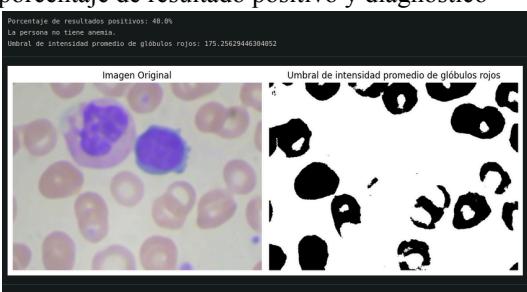
Erosión y transformada de Watershed



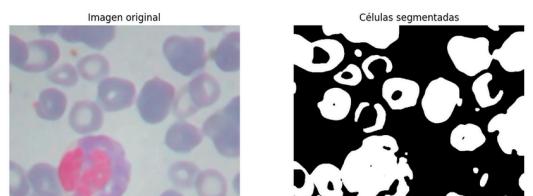
Mostrar etiquetas de tipos de células y conteo de las mismas



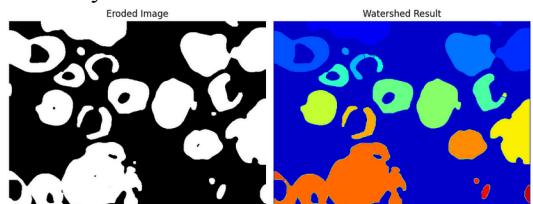
Umbral promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico



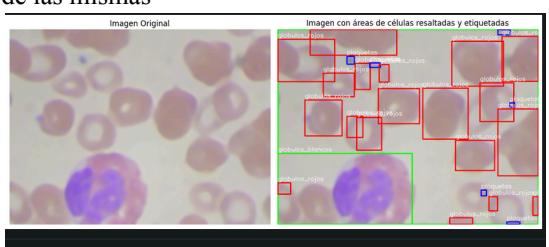
- Procesamiento tercera imagen  
Umbralización



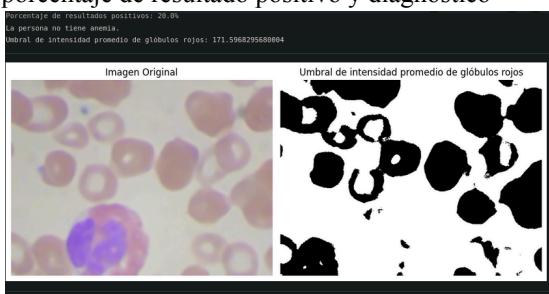
Erosión y transformada de Watershed



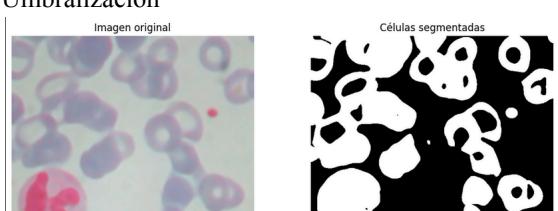
Mostrar etiquetas de tipos de células y conteo de las mismas



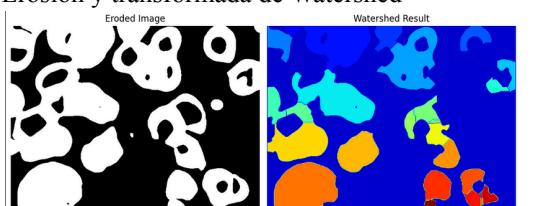
Umbral promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico



- Procesamiento cuarta imagen  
Umbralización



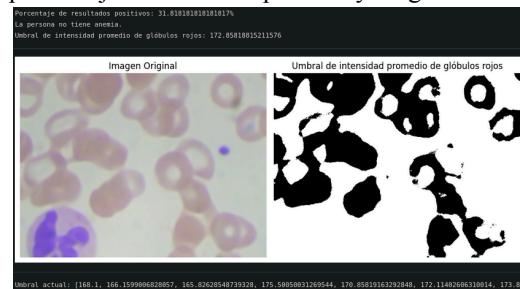
Erosión y transformada de Watershed



Mostrar etiquetas de tipos de células y conteo de las mismas



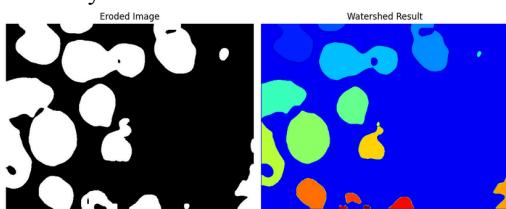
Umbral promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico



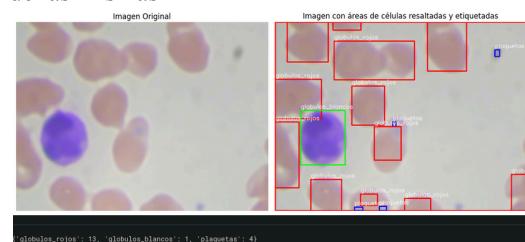
- Procesamiento quinta imagen  
Umbralización



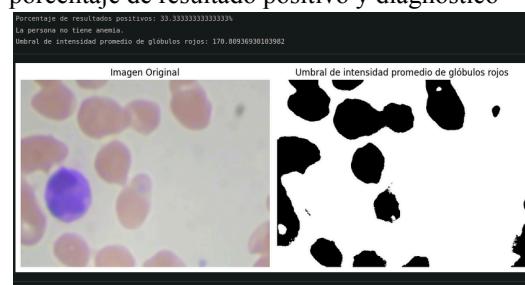
Erosión y transformada de Watershed



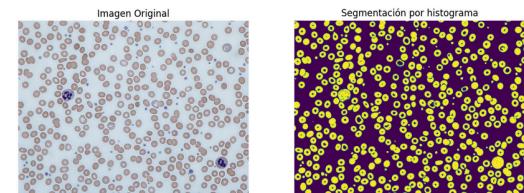
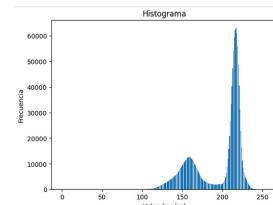
Mostrar etiquetas de tipos de células y conteo de las mismas



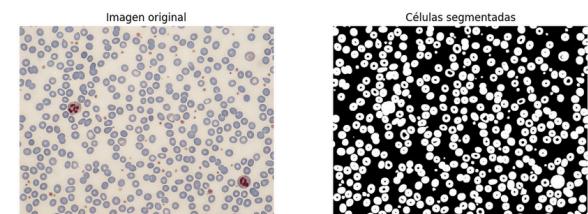
Umbral promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico



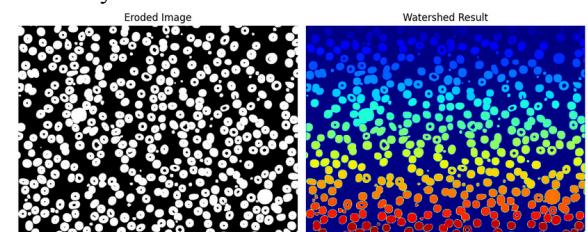
- Anemia Ferropénica
  - Procesamiento primera imagen  
Histograma de frecuencia



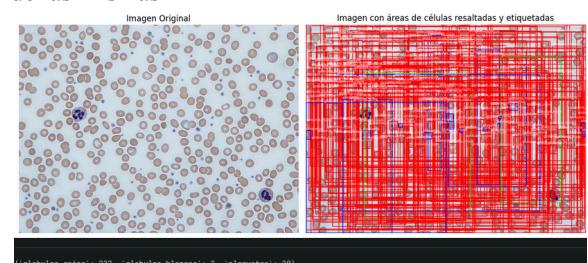
Umbralización



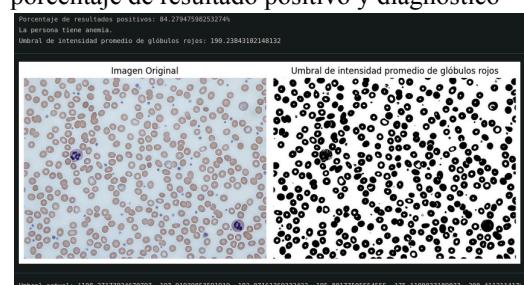
Erosión y transformada de Watershed



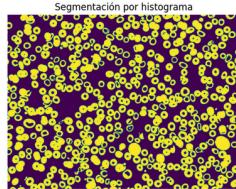
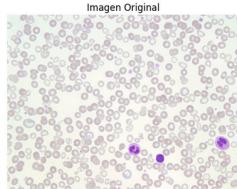
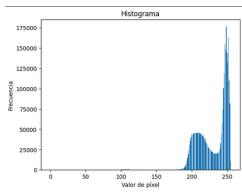
Mostrar etiquetas de tipos de células y conteo de las mismas



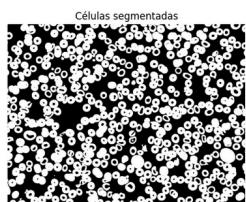
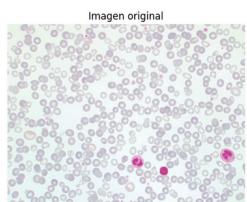
Umbral promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico



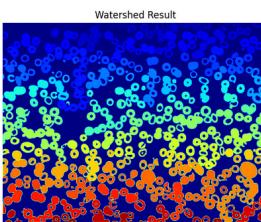
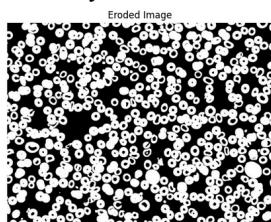
- Procesamiento segunda imagen  
Histograma de frecuencia



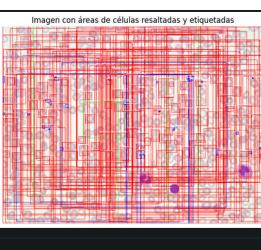
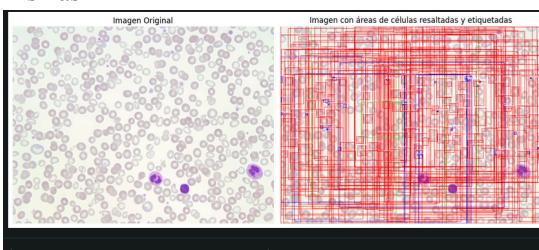
### Umbralización



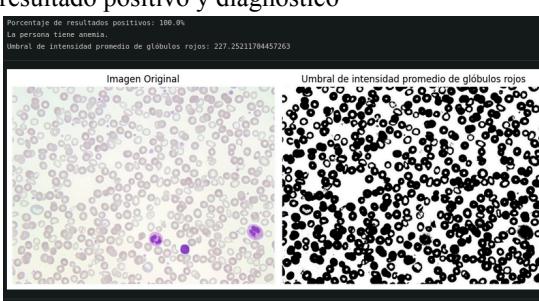
### Erosión y transformada de Watershed



Mostrar etiquetas de tipos de células y conteo de las mismas

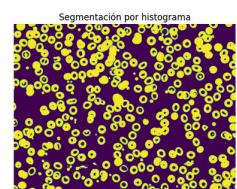
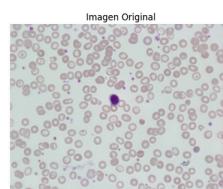
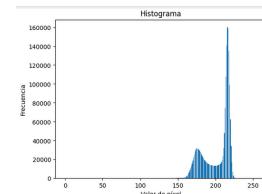


Umbrales promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico

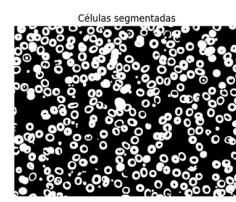
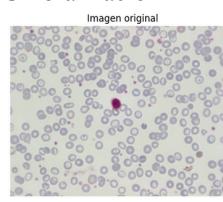


Umbrales actuales: [208, 7742828421809, 231, 35458896095093, 231, 292163558071, 230, 07929500398242, 229, 24453378165924, 232, 38616345]

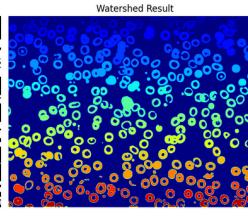
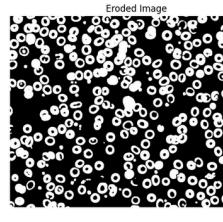
- Procesamiento tercera imagen  
Histograma de frecuencia



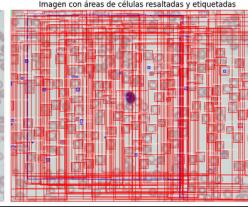
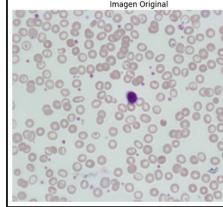
### Umbralización



### Erosión y transformada de Watershed



Mostrar etiquetas de tipos de células y conteo de las mismas



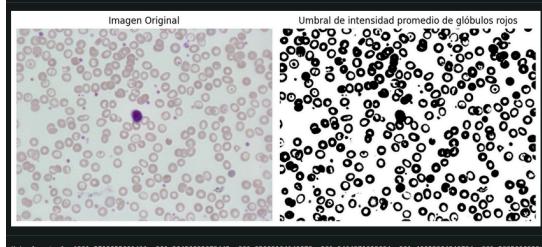
{ "glóbulos\_rojos": 227, "glóbulos\_blancos": 1, "plaquetas": 62 }

Umbrales promedio de glóbulos rojos, porcentaje de resultado positivo y diagnóstico

Porcentaje de resultados positivos: 97.797568281384%

La persona tiene anemia.

Umbrales de intensidad promedio de glóbulos rojos: 192.75473694327178



Umbrales actuales: [289, 5713655663411, 281, 76426892073445, 291, 05993184149878, 291, 3404271023684, 291, 40652864840734, 291, 58850869328]

### Conclusiones

- Este proyecto logró avances significativos en el reconocimiento automatizado de los principales tipos de células sanguíneas (glóbulos rojos, glóbulos blancos y plaquetas) a partir del procesamiento y análisis de imágenes microscópicas. Se implementaron con éxito técnicas de procesamiento de

imágenes como filtrado, segmentación, umbralización y clasificación, permitiendo la identificación y etiquetado de las células según su tipo.

- Los resultados obtenidos en la detección automatizada de enfermedades hematológicas como la anemia ferropénica a partir del análisis de imágenes de células sanguíneas son prometedores. Si bien sólo se evaluó preliminarmente la detección de anemia ferropénica, este trabajo sienta las bases para extender la metodología al diagnóstico de otros padecimientos sanguíneos.
- Este proyecto representa un avance en la aplicación de técnicas de procesamiento y análisis digital de imágenes al campo de la hematología. Los logros en la clasificación de células sanguíneas y la detección temprana de enfermedades a partir de imágenes microscópicas pueden tener un alto impacto al permitir diagnósticos más rápidos, precisos y automatizados.

## Referencias

- Shenggan. (s. f.). GitHub - Shenggan/BCCD\_Dataset: BCCD (Blood Cell Count and Detection) dataset is a small-scale dataset for blood cells detection. GitHub. [https://github.com/Shenggan/BCCD\\_Dataset](https://github.com/Shenggan/BCCD_Dataset)
- Kumar, V., Abbas, A. K., Aster, J. C., & Robbins, S. L. (2019). Patología estructural y funcional. Elsevier.
- Hoffbrand, A. V., Moss, P. A. H., & Pettit, J. E. (2016). Essential haematology. John Wiley & Sons.
- Gomez, D. M., et al. (2020). Automated image analysis for bone marrow plasma cell quantification: Can it replace manual counts? *Acta Histochemica*, 122(2), 151487.
- Johnson, L. M., et al. (2021). Automated quantification of red blood cell morphology on blood smears. *Journal of Clinical Pathology*, 74(1), 55-61.
- Smith, A. B., et al. (2018). Manual blood cell differential count. In StatPearls [Internet]. StatPearls Publishing.
- Brown, R. S., & Martinez, J. (2022). Advancements in automated blood cell morphology analysis. *American Journal of Clinical Pathology*, 158(3), 327-337.
- Smith, A. B., et al. (2018). Manual blood cell differential count. In StatPearls [Internet]. StatPearls Publishing.