



# PROYECTO\_ENTREGA1

Luis Mateo Hincapié Martínez  
1038417207

## Clasificación de tipos de células sanguíneas periféricas a partir de imágenes

El diagnóstico de enfermedades basadas en la sangre a menudo implica la identificación y caracterización de muestras de sangre de pacientes. Los métodos automatizados para detectar y clasificar los subtipos de células sanguíneas tienen importantes aplicaciones médicas.

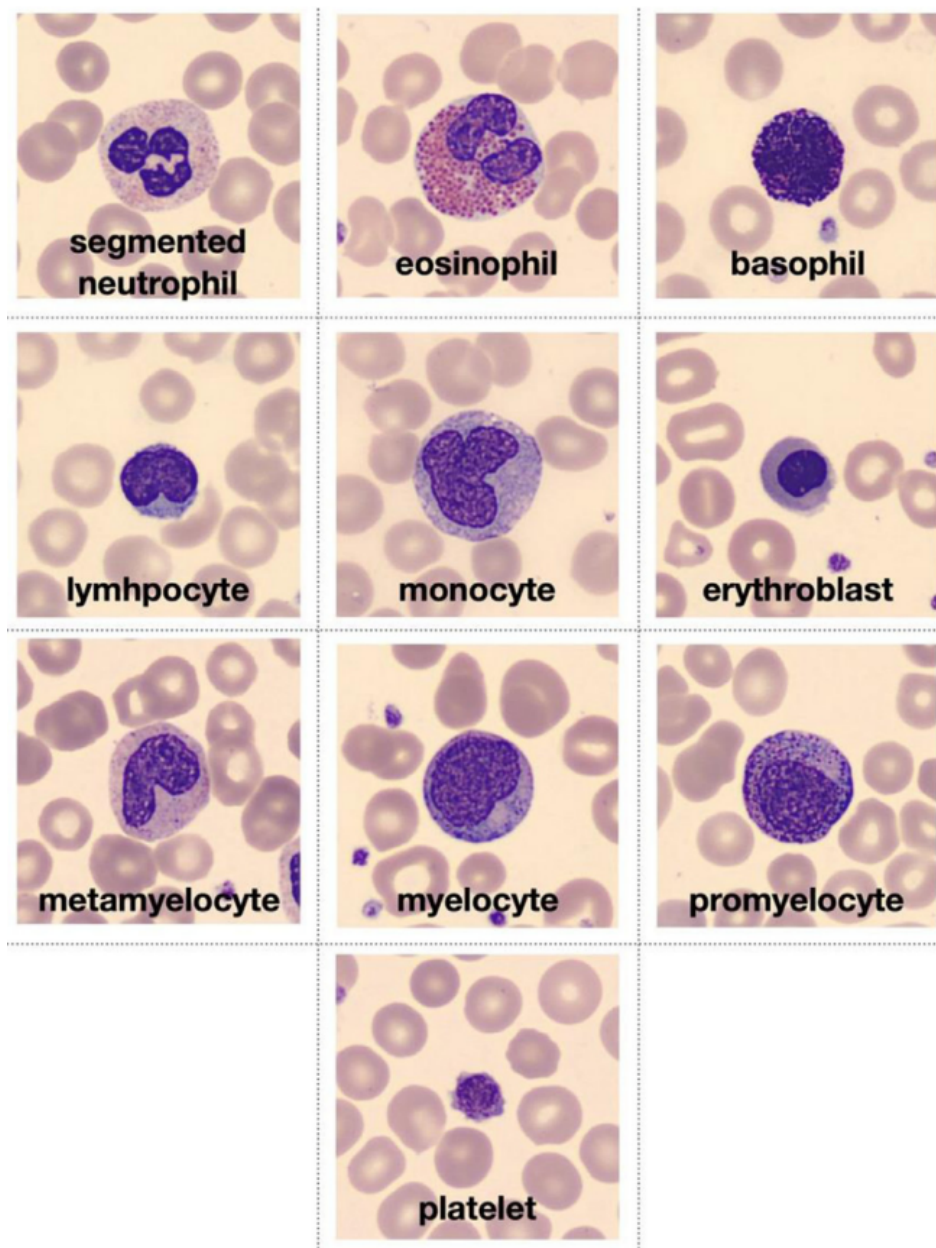
La propuesta es usar un conjunto de datos de ocho tipos de células de sangre periférica para clasificar nuevas muestras en el respectivo grupo de células, esto mediante el uso de redes neuronales.

### Conjunto de datos

Se obtuvo un conjunto de datos de 17092 imágenes de ocho clases de células sanguíneas periféricas normales utilizando el analizador CellaVision DM96. Posteriormente un grupo de patólogos clasificaron todas las imágenes en los siguientes tipos: neutrófilos, eosinófilos, basófilos, linfocitos, monocitos, granulocitos inmaduros (mielocitos, metamielocitos y promielocitos), eritroblastos y plaquetas.

El conjunto de datos se distribuye en 8 carpetas de la siguiente manera:

CELL TYPE	TOTAL OF IMAGES BY TYPE
neutrophils	3329
eosinophils	3117
basophils	1218
lymphocytes	1214
monocytes	1420
immature granulocytes (metamyelocytes, myelocytes and promyelocytes)	2895
erythroblasts	1551
platelets (thrombocytes)	2348
Total	17,092



*Figura 1. Ejemplo de los diferentes tipos de células.*

El conjunto de datos que se encuentra disponible en <https://data.mendeley.com/datasets/snkd93bnjr/1> fue compartido por el Hospital Clínico y Provincial de Barcelona desde el 8 de abril de 2020.

## Métricas de desempeño

La métrica principal que se desea optimizar será la precisión (accuracy) global del modelo, que se analizará en conjunto a la gráfica de curva ROC. En caso de obtener un puntaje considerable de *accuracy* con determinado modelo se usará el puntaje F1 o la matriz de confusión para un análisis más profundo.

En este caso no se está determinando una patología como tal, lo que se quiere es clasificar tipos de células, para luego ser contadas y determinar posibles patologías. Existen dilemas éticos y morales que hacen difícil dar un estimado de cual sería la precisión adecuada, por lo que la elección se realizó de forma más o menos arbitraria.

Para el problema que se abordara en este proyecto se buscará un accuracy del 95% y un F1 lo más grande posible, partiendo del hecho que tanto la reducción como el aumento en el número de células puede ocasionar diferentes patologías, por lo que no podemos “sesgar” el modelo para evitar que no descarte pacientes enfermos así se equivoque con los sanos como se hace en otras circunstancias, ya que el problema del proyecto no es de este tipo.

## **Criterio de desempeño deseable en producción**

En el ámbito médico se tienen ciertas tasas de errores que ya han sido ampliamente estudiadas y puestas a disposición de las entidades responsables de asegurar la calidad de mediciones cuantitativas en los laboratorios clínicos. Incluso cada grupo de células clasificadas tiene un porcentaje de tolerancia de error diferente. Actualmente desconozco los valores de cada uno de estos porcentajes, pero podría dar una estimación de que si el modelo logra permanecer dentro del rango de estos porcentajes es viable que salga a producción.

## **Referencias**

Acevedo, Andrea; Merino, Anna; Alferez, Santiago; Molina, Ángel; Boldú, Laura; Rodellar, José (2020), “A dataset for microscopic peripheral blood cell images for development of automatic recognition systems”, Mendeley Data, V1, doi: 10.17632/snkd93bnjr.1