

# Trabajo final de curso “Machine Learning e Imágenes en Python”

Automatización del conteo de unidades formadoras de colonias bacterianas

Benelli, Federico Ezequiel; Rovezzi, Juan Pablo

## Introducción

Las cuantificación de unidades formadoras de colonia (UFC) es una técnica empleada en microbiología con la finalidad de cuantificar el número de microorganismos viables (es decir, capaces de desarrollarse en un medio de cultivo bajo condiciones controladas) en una muestra de interés. Se utilizan principalmente en ámbitos bioquímicos y bromatológicos/alimenticios.

La técnica se desarrolla colocando una dilución de la muestra de interés sobre una placa de petri con un medio de cultivo que brinda nutrientes a los microorganismos de interés. Tras un tiempo de incubación dependiente del tipo de microorganismo a estudiar se realiza un conteo de las UFC presentes en la placa (Figura 1).

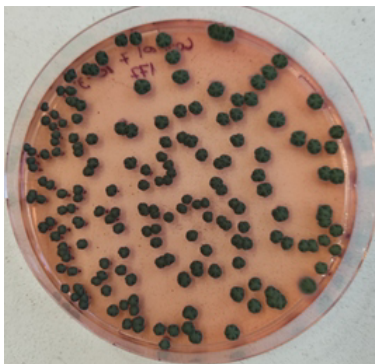


Figura 1: Cápsula de Petri

En laboratorios donde se realizan estudios de UFC de manera rutinaria y en cantidades, contabilizar colonias puede volverse una tarea exhaustiva y propensa a errores, en este trabajo se presenta una herramienta sencilla que permite automatizar esta actividad.

## Metodología

Para realizar el conteo de UFCs de una manera automatizada se utilizó un enfoque de entrenamiento semi-supervisado, utilizando un modelo K-Nearest Neighbors (KNN). La implementación fue realizada en Python, con soporte de las librerías opencv, matplotlib, pandas, numpy y sklearn.

Para llevarlo a cabo se requirieron 3 etapas:

- Carga de imagen y preprocesamiento
- Entrenamiento de modelo y categorización de datos
- Conteo de UFCs

## Carga de imagen y preprocesamiento

Partiendo de una imagen similar a la Figura 1, se selecciona un área que representa específicamente a la cápsula de petri, una vez seleccionada se extrae además una subárea correspondiente a una serie de colonias que se desea contabilizar. Posteriormente se realiza un recorte del área exterior a la cápsula, con el fin de facilitar la categorización posterior al tener una imagen más limpia.

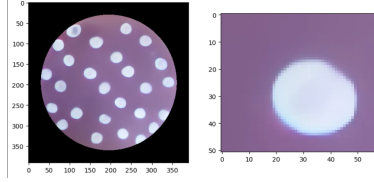


Figura 2: Imagen recortada

Una vez teniendo ambas imágenes listas, se procede a realizar un binarizado de la imagen de la subárea, para esto se ajustan thresholds inferior y superior hasta obtener la colonia aislada. Esto se realiza superponiendo la imagen original y la imagen binarizada con transparencia para asegurar un filtro lo más fidedigno posible.

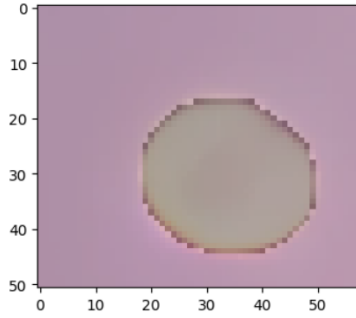


Figura 3: Recorte con threshold superpuesto

La imagen de subárea se convierte a un DataFrame y se utilizan los datos de la imagen binaria para categorizar a la colonia.

## Entrenamiento de modelo y categorización de datos totales

Con el set de datos semi-catalogado se entrena un modelo KNN, en el caso de estudio se realizaron distintas pruebas de números de vecinos a utilizar y se obtuvieron buenos resultados con 20 vecinos.

Una vez teniendo un modelo entrenado se realiza una categorización en todo el set de datos de la imagen. En la Figura 4 puede verse las colonias detectadas por KNN.

## Conteo de unidades formadoras de colonias

Para el conteo de UFCs se implementaron dos metodologías:

- Conteo por relación de áreas
- Conteo de puntos

### *Conteo por relación de áreas*

Sabiendo el número de colonias en la subimagen realizada se calcula el área promedio que corresponde a una colonia. El área se mide en unidades de pixeles y el área promedio se determina según:

$$A_{promedio} = \frac{N_{Pixeles}}{N_{UFCs}}$$

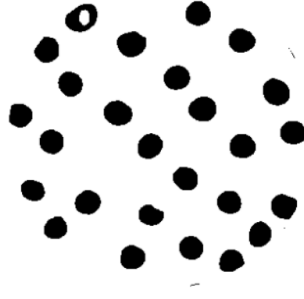


Figura 4: Cluster detectado por KNN

Una vez determinada el área promedio de las UFCs se procede a calcular el número de colonias. lo cual se realiza mediante la relación:

$$N_{UFCs \text{ totales}} = \frac{N_{\text{Píxeles}}}{A_{\text{promedio}}}$$

Lo que corresponde al cociente entre el número total de píxeles que ocupan los datos catalogados como UFC de interés y el área promedio determinada previamente.

#### *Conteo de puntos*

Como método alternativo se utilizó un algoritmo contador de componentes (`cv2.connectedComponents`) utilizando la imagen de las colonias ya catalogadas mediante KNN.

## Conclusiones y perspectivas futuras

Se pudo implementar un modelo KNN capaz de identificar y contabilizar de manera eficiente y en gran medida automatizada unidades formadoras de colonias. Como extensión de este trabajo podría realizarse el conteo de múltiples colonias distintas dentro de la misma cápsula. Además, se podrían realizar mejoras en la técnica de categorización, ya que en algunos casos brillos o algún tipo de ruido particular puede ocasionar problemas en la detección.