

**Título:** Clasificador de hongos: ¿Qué especie observamos?

**Autor:** Dulce Margarita Sánchez López

**Materias:** Aprendizaje Automático Supervisado (AAS) y Adquisición y Tratamiento de la Información(ATI).

En el ámbito de la biología y la ecología, la correcta identificación de especies de hongos resulta fundamental tanto para la conservación ambiental como para la seguridad alimentaria. Existen miles de especies con morfologías similares donde a la hora de selección puede ser un tanto peligroso para salir de campo y se topan con especies que pueden ser confundidas como comestibles.

Es por ello que por mi pasión sobre el mundo de la vida y mi carrera aplicada es que decidí hacer un clasificador que me permita juntar mis conocimientos en algo que podría serle útil a más personas que se encuentren con dificultades en clasificación.

El conjunto de datos proviene de Kaggle bajo el nombre de “Mushroom species recognition” cuyo peso en imágenes es de alrededor de 12.4 GB; este dataset contiene imágenes RGB clasificadas por especie, cada carpeta representando una clase distinta de hongo. Las imágenes presentan variaciones en color, textura, forma y entorno, lo cual favorece la generalización del modelo. El tipo de datos utilizado será imágenes, con etiquetas categóricas que corresponden a las especies presentes en el conjunto.

Para lograr el entrenamiento, por parte de la materia de ATI someteré las imágenes a un proceso de adquisición, limpieza y estandarización. Las transformaciones incluirán: redimensionamiento uniforme ( $256 \times 256$  px) sin distorsionar la forma, aplicando una técnica de relleno reflejado (letterbox) para conservar proporciones; normalización de valores de píxeles (0–1) para estabilizar el aprendizaje; y aumento de datos mediante rotaciones, volteos horizontales y cambios leves de brillo, con el fin de incrementar la variabilidad y reducir el sobreajuste.

Cabe destacar que no se usarán todas las imágenes, se hará una selección de las mismas disminuyendo la cantidad de especies y logrando bajar el uso de memoria para evitar colisiones en la misma.

Antes del entrenamiento, se realizará un análisis exploratorio del conjunto de imágenes. Se mostrarán distribuciones del número de imágenes por clase, ejemplos visuales de cada especie, comparaciones entre histogramas de color o textura y ejemplos antes y después del preprocesamiento. Estas visualizaciones permitirán detectar sesgos, desequilibrios o similitudes morfológicas entre especies, y justificar las decisiones de transformación de datos.

El modelo principal será una Red Neuronal Convolutacional (CNN) basada en Transfer Learning, utilizando una arquitectura preentrenada como MobileNetV2 o ResNet50, debido a su eficiencia para el reconocimiento de patrones visuales. Se entrenará con división de datos en entrenamiento (80%), validación (10%) y prueba (10%), evaluando el desempeño mediante precisión (accuracy), F1-score y matriz de confusión. Adicionalmente, se emplearán modelos de aprendizaje automático como Random Forest, Gaussian Naive Bayes, AdaBoost y Boosting.

Se espera obtener un clasificador completo que nos permita dar ese plus esperado; cuya hipótesis es poder identificar en una imagen o video a tiempo real que especie de hongo

es. Usando el sistema de interfaz alternativa (Gradio o Streamlit). Pero antes de ello poder hacer funcionar el clasificador de manera óptima y obtener resultados favorables para cumplir con los requerimientos establecidos en la rúbrica de ambas materias.