```
# Manejo de archivos y sistema operativo
import os  # Interacción con el sistema operativo (rutas, archivos, etc.)
import itertools  # Herramientas para crear combinaciones, permutaciones, etc.
from pathlib import Path  # Manejo de rutas de archivos de forma más robusta y orientada a objetos

# # Manejo de datos
# import numpy as np  # Computación numérica eficiente con arrays
import pandas as pd  # Manipulación y análisis de datos en estructuras tipo DataFrame
# import polars as pl  # Alternativa a pandas, más rápida para grandes volúmenes de datos

# # Clasificación de imágenes
# from keras.models import Sequential, Model
# from keras.layers import Conv2D, MaxPool2D, Dense, Flatten, Dropout, BatchNormalization, Input
# from keras.optimizers import Adam
# from keras.callbacks import TensorBoard, ModelCheckpoint
# from keras.utils import to_categorical
# from keras.preprocessing import image
# from keras.applications.imagenet_utils import preprocess_input, decode_predictions
# from keras.applications.vgg16 import VGG16
# from tensorflow.keras.preprocessing.image import ImageDataGenerator
# from sklearn.utils import shuffle
# from sklearn.model_selection import train_test_split
# import cv2
# import matplotlib.pyplot as plt
# from tensorflow.keras.preprocessing import image_dataset_from_directory
# %matplotlib inline

# # Manejo de advertencias
# import warnings  # Permite controlar las advertencias del sistema
# warnings.filterwarnings("ignore")  # Suprime todas las advertencias
```

```
#Conectar con google drive
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=Tr
```

Leer CSV

```
BCN20000=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/bcn20000_metadata_2025-06-19.csv')
Challenge2018=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2018-task-3-training_metadata_2025-06-22.csv')
Challenge2024=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2024-training_metadata_2025-06-22.csv')
Challenge2016_train=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2016-training_metadata_2025-06-22.csv')
Challenge2016_test=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2016-test_metadata_2025-06-22.csv')
HAM10000=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/ham10000_metadata_2025-06-26.csv')
Challenge2020_train=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2020-training_metadata_2025-06-26.csv')
Challenge2020_test=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2020-test_metadata_2025-06-26.csv')
```

```
/tmp/ipython-input-3054675085.py:3: DtypeWarning: Columns (12,16) have mixed types. Specify dtype option on import or set lo
  Challenge2024=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2024-training_metadata_2025-06-22.csv')
/tmp/ipython-input-3054675085.py:7: DtypeWarning: Columns (14,20,22) have mixed types. Specify dtype option on import or set
  Challenge2020_train=pd.read_csv('/content/drive/MyDrive/Metadatos/CSV/challenge-2020-training_metadata_2025-06-26.csv')
```

```
archivos = {
    "BCN20000": "/content/drive/MyDrive/Metadatos/CSV/bcn20000_metadata_2025-06-19.csv",
    "Challenge2018": "/content/drive/MyDrive/Metadatos/CSV/challenge-2018-task-3-training_metadata_2025-06-22.csv",
    "Challenge2024": "/content/drive/MyDrive/Metadatos/CSV/challenge-2024-training_metadata_2025-06-22.csv",
    "Challenge2016_train": "/content/drive/MyDrive/Metadatos/CSV/challenge-2016-training_metadata_2025-06-22.csv",
    "Challenge2016_test": "/content/drive/MyDrive/Metadatos/CSV/challenge-2016-test_metadata_2025-06-22.csv",
    "HAM10000": "/content/drive/MyDrive/Metadatos/CSV/ham10000_metadata_2025-06-26.csv",
    "Challenge2020_train": "/content/drive/MyDrive/Metadatos/CSV/challenge-2020-training_metadata_2025-06-26.csv",
    "Challenge2020_test": "/content/drive/MyDrive/Metadatos/CSV/challenge-2020-test_metadata_2025-06-26.csv"
}

# Recorrer los archivos y contar valores de 'diagnosis_3'
for nombre, ruta in archivos.items():
    df = pd.read_csv(ruta)

    if "diagnosis_3" in df.columns:
        print(f"\n {nombre} — Recuento de 'diagnosis_3':")
        print(df["diagnosis_3"].value_counts())
    else:
        print(f"\n {nombre} no tiene la columna 'diagnosis_3'")
```

```

 BCN20000 — Recuento de 'diagnosis_3':
diagnosis_3
```

```
Nevus                          5647
Melanoma, NOS                  4003
Basal cell carcinoma           3676
Seborrheic keratosis           1268
Solar or actinic keratosis     1088
Melanoma metastasis             633
Squamous cell carcinoma, NOS    559
Scar                            314
Solar lentigo                   283
Dermatofibroma                  168
Name: count, dtype: int64

 Challenge2018 — Recuento de 'diagnosis_3':
diagnosis_3
Nevus                          6705
Melanoma, NOS                  1113
Pigmented benign keratosis     1099
Basal cell carcinoma            514
Squamous cell carcinoma, NOS    197
Solar or actinic keratosis      130
Dermatofibroma                  115
Name: count, dtype: int64
/tmp/ipython-input-627809016.py:14: DtypeWarning: Columns (12,16) have mixed types. Specify dtype option on import or set
  df = pd.read_csv(ruta)

 Challenge2024 — Recuento de 'diagnosis_3':
diagnosis_3
Nevus                                             443
Basal cell carcinoma                              163
Melanoma in situ                                   80
Atypical melanocytic neoplasm                      64
Melanoma Invasive                                  63
Seborrheic keratosis                               57
Squamous cell carcinoma in situ                    49
Solar or actinic keratosis                         39
Squamous cell carcinoma, Invasive                  22
Melanoma, NOS                                      13
Solar lentigo                                      12
Lichen planus like keratosis                       11
Dermatofibroma                                     11
Atypical intraepithelial melanocytic proliferation 11
Verruca                                             7
Lentigo NOS                                         5
Pigmented benign keratosis                          3
Hemangioma                                          3
Angiofibroma                                        2
Squamous cell carcinoma, NOS                        2
Trichilemmal or isthmic-catagen or pilar cyst       1
Melanoma metastasis                                 1
Scar                                                1
Hidradenoma                                         1
Fibroepithelial polyp                               1
Name: count, dtype: int64
```

Cambio de nombres

```
#Leer CSV de los Diagnosis
diagnosis=pd.read_csv('/content/drive/MyDrive/Metadatos/cambio de nombres - Hoja 1.csv')
print(diagnosis)
```

```
                      Clasificacion             Nombre final
0                             Nevus                     Nevo
1                      Melanoma, NOS     Dermatosis cancerosa
2       Squamous cell carcinoma, NOS     Dermatosis cancerosa
3                      Solar lentigo           Otras lesiones
4               Basal cell carcinoma     Dermatosis cancerosa
..                               ...                      ...
96                  Lentigo simplex           Otras lesiones
97      Lichen planus like keratosis           Otras lesiones
98        Solar or actinic keratosis  Dermatosis precancerosa
99                              Scar           Otras lesiones
100  Squamous cell carcinoma in situ     Dermatosis cancerosa

[101 rows x 2 columns]
```

```
#Eliminar duplicados según la columna "Clasificación"
diagnosis_unicos=diagnosis.drop_duplicates(subset='Clasificacion')
print(diagnosis_unicos)
```

```
                                  Clasificacion             Nombre final
0                                         Nevus                     Nevo
1                                  Melanoma, NOS     Dermatosis cancerosa
2                   Squamous cell carcinoma, NOS     Dermatosis cancerosa
3                                  Solar lentigo           Otras lesiones
4                           Basal cell carcinoma     Dermatosis cancerosa
```

```
5                              Melanoma metastasis     Dermatosis cancerosa
6                            Seborrheic keratosis         Tumores benignos
7                      Solar or actinic keratosis  Dermatosis precancerosa
8                                  Dermatofibroma         Tumores benignos
9                                            Scar           Otras lesiones
12                     Pigmented benign keratosis         Tumores benignos
19                                Melanoma in situ     Dermatosis cancerosa
20                  Atypical melanocytic neoplasm  Dermatosis precancerosa
21                                Melanoma Invasive     Dermatosis cancerosa
23                 Squamous cell carcinoma in situ     Dermatosis cancerosa
25               Squamous cell carcinoma, Invasive     Dermatosis cancerosa
28                   Lichen planus like keratosis           Otras lesiones
30  Atypical intraepithelial melanocytic prolifera...  Dermatosis precancerosa
31                                         Verruca           Otras lesiones
32                                     Lentigo NOS           Otras lesiones
34                                      Hemangioma         Tumores benignos
35                                    Angiofibroma         Tumores benignos
37     Trichilemmal or isthmic-catagen or pilar cyst         Tumores benignos
40                                     Hidradenoma         Tumores benignos
41                             Fibroepithelial polyp         Tumores benignos
48                                  Lentigo simplex           Otras lesiones
50                                  Ink-spot lentigo           Otras lesiones
56                                   Lentigo simplex           Otras lesiones
60                                   Epidermal nevus                    Nevo
63                             Mucosal melanotic macule           Otras lesiones
82                                    Warty dyskeratoma           Otras lesiones
83                          Cafe au lait macule or patch           Otras lesiones
85                                       Tricholemmoma         Tumores benignos
86                                       Porokeratosis           Otras lesiones
87                                Large cell acanthoma         Tumores benignos
```

```
#Crear diccionario de traducción
traduccion=dict(zip(diagnosis_unicos['Clasificacion'],diagnosis_unicos['Nombre final']))
print("Diccionario de traducción:")
print(traduccion)
```

```
Diccionario de traducción:
{'Nevus': 'Nevo', 'Melanoma, NOS': 'Dermatosis cancerosa', 'Squamous cell carcinoma, NOS': 'Dermatosis cancerosa', 'Solar le
```

```
#Reemplazo
Challenge2018['diagnosis_3'] = Challenge2018['diagnosis_3'].replace(traduccion)
Challenge2024['diagnosis_3'] = Challenge2024['diagnosis_3'].replace(traduccion)
Challenge2016_train['diagnosis_3'] = Challenge2016_train['diagnosis_3'].replace(traduccion)
Challenge2016_test['diagnosis_3'] = Challenge2016_test['diagnosis_3'].replace(traduccion)
HAM10000['diagnosis_3'] = HAM10000['diagnosis_3'].replace(traduccion)
Challenge2020_train['diagnosis_3'] = Challenge2020_train['diagnosis_3'].replace(traduccion)
Challenge2020_test['diagnosis_3'] = Challenge2020_test['diagnosis_3'].replace(traduccion)
BCN20000['diagnosis_3'] = BCN20000['diagnosis_3'].replace(traduccion)
```

```
#Challenge2018.head()
```

```
Challenge2018.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10015 entries, 0 to 10014
Data columns (total 15 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   isic_id                 10015 non-null  object
 1   attribution             10015 non-null  object
 2   copyright_license       10015 non-null  object
 3   age_approx              9921 non-null   float64
 4   anatom_site_general     8377 non-null   object
 5   anatom_site_special     464 non-null    object
 6   concomitant_biopsy      10015 non-null  bool
 7   diagnosis_1             10015 non-null  object
 8   diagnosis_2             10015 non-null  object
 9   diagnosis_3             9873 non-null   object
 10  diagnosis_confirm_type  10015 non-null  object
 11  image_type              10015 non-null  object
 12  lesion_id               10015 non-null  object
 13  melanocytic             10015 non-null  bool
 14  sex                     9968 non-null   object
dtypes: bool(2), float64(1), object(12)
memory usage: 1.0+ MB
```

```
valores_unicos1 = Challenge2018['diagnosis_3'].unique()
print("Challenge2018:",valores_unicos1)
valores_unicos2 = Challenge2024['diagnosis_3'].unique()
print("Challenge2024:",valores_unicos2)
valores_unicos3 = Challenge2016_train['diagnosis_3'].unique()
print("Challenge2016_train:",valores_unicos3)
valores_unicos4 = Challenge2016_test['diagnosis_3'].unique()
```

```
print("Challenge2016_test:",valores_unicos4)
valores_unicos5 = HAM10000['diagnosis_3'].unique()
print("HAM10000:",valores_unicos5)
valores_unicos6 = Challenge2020_train['diagnosis_3'].unique()
print("Challenge2020_train:",valores_unicos6)
valores_unicos7 = Challenge2020_test['diagnosis_3'].unique()
print("Challenge2020_test:",valores_unicos7)
valores_unicos8 = BCN20000['diagnosis_3'].unique()
print("BCN20000:",valores_unicos8)
```

```
Challenge2018: ['Nevo' 'Dermatosis cancerosa' 'Tumores benignos' nan
 'Dermatosis precancerosa']
Challenge2024: [nan 'Otras lesiones' 'Nevo' 'Dermatosis cancerosa'
 'Dermatosis precancerosa' 'Tumores benignos']
Challenge2016_train: ['Nevo' 'Dermatosis cancerosa' nan 'Otras lesiones'
 'Dermatosis precancerosa' 'Tumores benignos']
Challenge2016_test: ['Nevo' 'Dermatosis cancerosa' nan 'Otras lesiones'
 'Dermatosis precancerosa']
HAM10000: ['Nevo' 'Dermatosis cancerosa' 'Tumores benignos' nan
 'Dermatosis precancerosa']
Challenge2020_train: [nan 'Nevo' 'Dermatosis cancerosa' 'Tumores benignos' 'Otras lesiones'
 'Dermatosis precancerosa']
Challenge2020_test: [nan 'Nevo' 'Dermatosis cancerosa' 'Otras lesiones' 'Tumores benignos'
 'Dermatosis precancerosa']
BCN20000: ['Nevo' 'Dermatosis cancerosa' nan 'Otras lesiones' 'Tumores benignos'
 'Dermatosis precancerosa']
```

```
#Elimino las columas donde el valor de diagnosis_3 este vacio (=nan)
Challenge2018 = Challenge2018.dropna(subset=['diagnosis_3'])
Challenge2024 = Challenge2024.dropna(subset=['diagnosis_3'])
Challenge2016_train = Challenge2016_train.dropna(subset=['diagnosis_3'])
Challenge2016_test = Challenge2016_test.dropna(subset=['diagnosis_3'])
HAM10000 = HAM10000.dropna(subset=['diagnosis_3'])
Challenge2020_train = Challenge2020_train.dropna(subset=['diagnosis_3'])
Challenge2020_test = Challenge2020_test.dropna(subset=['diagnosis_3'])
BCN20000 = BCN20000.dropna(subset=['diagnosis_3'])
```

```
Challenge2018.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 9873 entries, 0 to 10014
Data columns (total 15 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   isic_id                 9873 non-null   object
 1   attribution             9873 non-null   object
 2   copyright_license       9873 non-null   object
 3   age_approx              9783 non-null   float64
 4   anatom_site_general     8266 non-null   object
 5   anatom_site_special     459 non-null    object
 6   concomitant_biopsy      9873 non-null   bool
 7   diagnosis_1             9873 non-null   object
 8   diagnosis_2             9873 non-null   object
 9   diagnosis_3             9873 non-null   object
 10  diagnosis_confirm_type  9873 non-null   object
 11  image_type              9873 non-null   object
 12  lesion_id               9873 non-null   object
 13  melanocytic             9873 non-null   bool
 14  sex                     9826 non-null   object
dtypes: bool(2), float64(1), object(12)
memory usage: 1.1+ MB
```

```
valores_unicos1 = Challenge2018['diagnosis_3'].unique()
print("Challenge2018:",valores_unicos1)
valores_unicos2 = Challenge2024['diagnosis_3'].unique()
print("Challenge2024:",valores_unicos2)
valores_unicos3 = Challenge2016_train['diagnosis_3'].unique()
print("Challenge2016_train:",valores_unicos3)
valores_unicos4 = Challenge2016_test['diagnosis_3'].unique()
print("Challenge2016_test:",valores_unicos4)
valores_unicos5 = HAM10000['diagnosis_3'].unique()
print("HAM10000:",valores_unicos5)
valores_unicos6 = Challenge2020_train['diagnosis_3'].unique()
print("Challenge2020_train:",valores_unicos6)
valores_unicos7 = Challenge2020_test['diagnosis_3'].unique()
print("Challenge2020_test:",valores_unicos7)
valores_unicos8 = BCN20000['diagnosis_3'].unique()
print("BCN20000:",valores_unicos8)
```

```
Challenge2018: ['Nevo' 'Dermatosis cancerosa' 'Tumores benignos'
 'Dermatosis precancerosa']
Challenge2024: ['Otras lesiones' 'Nevo' 'Dermatosis cancerosa' 'Dermatosis precancerosa'
 'Tumores benignos']
Challenge2016_train: ['Nevo' 'Dermatosis cancerosa' 'Otras lesiones' 'Dermatosis precancerosa'
```

```
 'Tumores benignos']
Challenge2016_test: ['Nevo' 'Dermatosis cancerosa' 'Otras lesiones' 'Dermatosis precancerosa']
HAM10000: ['Nevo' 'Dermatosis cancerosa' 'Tumores benignos'
 'Dermatosis precancerosa']
Challenge2020_train: ['Nevo' 'Dermatosis cancerosa' 'Tumores benignos' 'Otras lesiones'
 'Dermatosis precancerosa']
Challenge2020_test: ['Nevo' 'Dermatosis cancerosa' 'Otras lesiones' 'Tumores benignos'
 'Dermatosis precancerosa']
BCN20000: ['Nevo' 'Dermatosis cancerosa' 'Otras lesiones' 'Tumores benignos'
 'Dermatosis precancerosa']
```

Comprobamos que se eliminaron las filas vacías de la columna nan

Guardar los nuevos CSV con las categorias finales

```
ruta_destino='/content/drive/MyDrive/Metadatos/CSVs limpios'
```

```
Challenge2018.to_csv(os.path.join(ruta_destino, 'Challenge2018_categoriasfinales.csv'), index=False)
Challenge2024.to_csv(os.path.join(ruta_destino, 'Challenge2024_categoriasfinales.csv'), index=False)
Challenge2016_train.to_csv(os.path.join(ruta_destino, 'Challenge2016_train_categoriasfinales.csv'), index=False)
Challenge2016_test.to_csv(os.path.join(ruta_destino, 'Challenge2016_test_categoriasfinales.csv'), index=False)
Challenge2020_train.to_csv(os.path.join(ruta_destino, 'Challenge2020_train_categoriasfinales.csv'), index=False)
Challenge2020_test.to_csv(os.path.join(ruta_destino, 'Challenge2020_test_categoriasfinales.csv'), index=False)
HAM10000.to_csv(os.path.join(ruta_destino, 'HAM10000_categoriasfinales.csv'), index=False)
BCN20000.to_csv(os.path.join(ruta_destino, 'BCN20000_categoriasfinales.csv'), index=False)
```

Metadata 2024 - Quitando los nevos

```
Challenge2024=pd.read_csv('/content/drive/MyDrive/Metadatos/CSVs limpios/Challenge2024_categoriasfinales.csv')
```

```
valores_unicoss = Challenge2024['diagnosis_3'].unique()
print("Challenge2024",valores_unicoss)
```

```
Challenge2024 ['Otras lesiones' 'Nevo' 'Dermatosis cancerosa' 'Dermatosis precancerosa'
 'Tumores benignos']
```

```
# Eliminar filas donde diagnosis_3 sea "Nevo"
Challenge2024 = Challenge2024[Challenge2024['diagnosis_3'] != 'Nevo']
```

```
valores_unicoss = Challenge2024['diagnosis_3'].unique()
print("Challenge2024",valores_unicoss)
```

```
Challenge2024 ['Otras lesiones' 'Dermatosis cancerosa' 'Dermatosis precancerosa'
 'Tumores benignos']
```

```
ruta_destino='/content/drive/MyDrive/Metadatos/CSVs limpios'
```

```
Challenge2024.to_csv(os.path.join(ruta_destino, 'Challenge2024_categoriasfinalesNONEVO.csv'), index=False)
```

```
Prueba=pd.read_csv('/content/drive/MyDrive/Metadatos/CSVs limpios/Challenge2024_categoriasfinalesNONEVO.csv')
print(Prueba['diagnosis_3'].unique())
```

```
['Otras lesiones' 'Dermatosis cancerosa' 'Dermatosis precancerosa'
 'Tumores benignos']
```

Elimino la clasificación Nevo del Challenge 2024 debido a que ya tenemos muchos ejemplos de Nevo, intentando "balancear" las clases