Análisis de resonancias magnéticas de tumores cerebrales

Francisco Javier Mercader Martínez

1 Importación de las librerías necesarias

```
import os
from PIL import Image
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from glob import glob
import random
# ------------------------------
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
# ------------------------------
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, Flatten
from tensorflow.keras.optimizers import Adamax
from tensorflow.keras.metrics import Precision, Recall
from tensorflow.keras.preprocessing.image import ImageDataGenerator
# ------------------------------
import warnings
warnings.filterwarnings('ignore')
```

2 Preprocesamiento

2.1 Cargar los datos

```
def train_df(tr_path, limit=100):
    classes, class_path = [], []
    for label in os.listdir(tr_path):
        label_path = os.path.join(tr_path, label)
        if os.path.isdir(label_path):
            images = os.listdir(label_path)
            sampled_images = random.sample(images, min(limit, len(images))) # Selecci
                ón aleatoria
            for image in sampled_images:
                classes.append(label)
                class_path.append(os.path.join(label_path, image))

    tr_df = pd.DataFrame({'Class Path': class_path, 'Class': classes})
    return tr_df
```

```
def test_df(tr_path, limit=100):
    classes, class_path = [], []
    for label in os.listdir(tr_path):
        label_path = os.path.join(tr_path, label)
        if os.path.isdir(label_path):
            images = os.listdir(label_path)
            sampled_images = random.sample(images, min(limit, len(images))) # Selecci
                ón aleatoria
```

```
            for image in sampled_images:
                classes.append(label)
                class_path.append(os.path.join(label_path, image))

    ts_df = pd.DataFrame({'Class Path': class_path, 'Class': classes})
    return ts_df
```

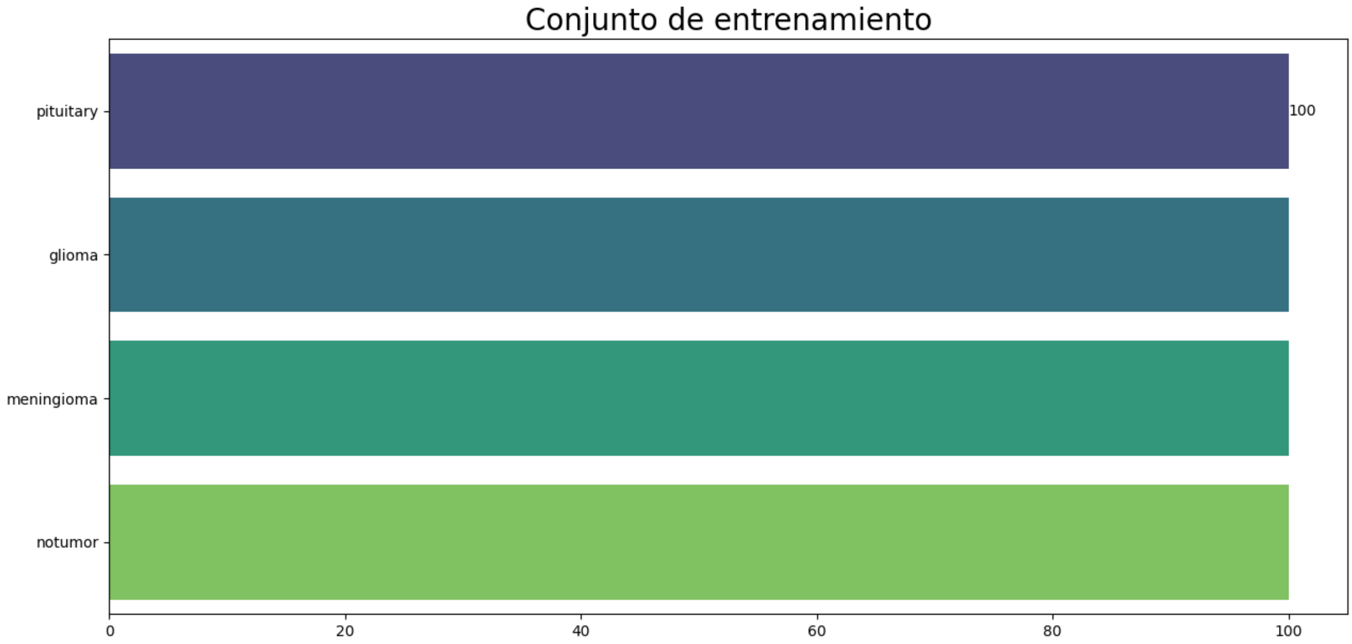
```
tr_df = train_df('kaggle/input/brain-tumor-mri-dataset/1/Training')

tr_df
```

	Class Path	Class
0	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	pituitary
1	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	pituitary
2	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	pituitary
3	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	pituitary
4	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	pituitary
...	...	...
395	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	notumor
396	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	notumor
397	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	notumor
398	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	notumor
399	kaggle/input/brain-tumor-mri-dataset/1/Trainin...	notumor

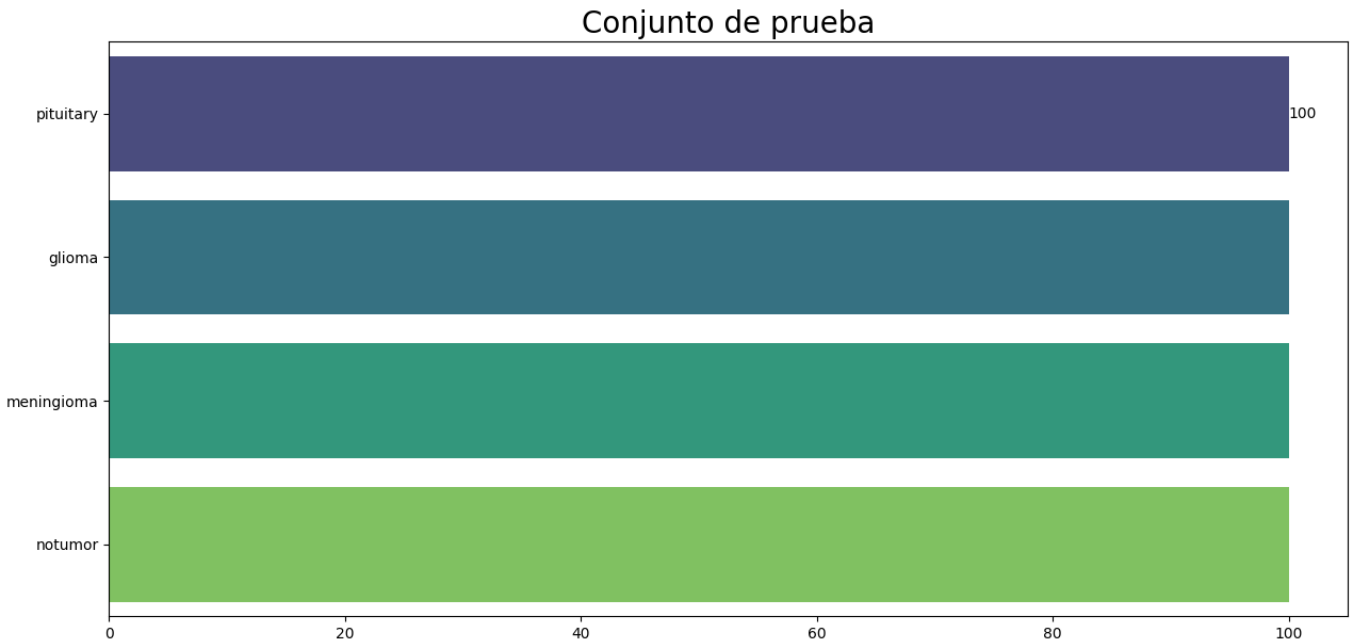
```
ts_df = test_df('kaggle/input/brain-tumor-mri-dataset/1/Testing')

ts_df
```

	Class Path	Class
0	kaggle/input/brain-tumor-mri-dataset/1/Testing...	pituitary
1	kaggle/input/brain-tumor-mri-dataset/1/Testing...	pituitary
2	kaggle/input/brain-tumor-mri-dataset/1/Testing...	pituitary
3	kaggle/input/brain-tumor-mri-dataset/1/Testing...	pituitary
4	kaggle/input/brain-tumor-mri-dataset/1/Testing...	pituitary
...	...	...
395	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor
396	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor
397	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor
398	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor
399	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor

```
# Contamos las imágenes en cada clase de los datos de entrenamiento
plt.figure(figsize=(15, 7))

ax = sns.countplot(data=tr_df, y=tr_df['Class'], palette='viridis')
plt.xlabel('')
plt.ylabel('')
plt.title("Conjunto de entrenamiento", fontsize=20)
ax.bar_label(ax.containers[0])
plt.show()
```




```
# Contamos las imágenes en cada clase de los datos de test
plt.figure(figsize=(15, 7))
ax = sns.countplot(y=ts_df['Class'], palette='viridis')

ax.set(xlabel='', ylabel='')
ax.bar_label(ax.containers[0])
plt.title('Conjunto de prueba', fontsize=20)
plt.show()
```



2.2 División de los datos

```
valid_df, ts_df = train_test_split(ts_df, train_size=0.5, random_state=20, stratify=
    ts_df['Class'])

valid_df
```

	Class Path	Class
193	kaggle/input/brain-tumor-mri-dataset/1/Testing...	glioma

	Class Path	Class
314	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor
99	kaggle/input/brain-tumor-mri-dataset/1/Testing...	pituitary
60	kaggle/input/brain-tumor-mri-dataset/1/Testing...	pituitary
148	kaggle/input/brain-tumor-mri-dataset/1/Testing...	glioma
...	...	...
135	kaggle/input/brain-tumor-mri-dataset/1/Testing...	glioma
308	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor
283	kaggle/input/brain-tumor-mri-dataset/1/Testing...	meningioma
312	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor
342	kaggle/input/brain-tumor-mri-dataset/1/Testing...	notumor

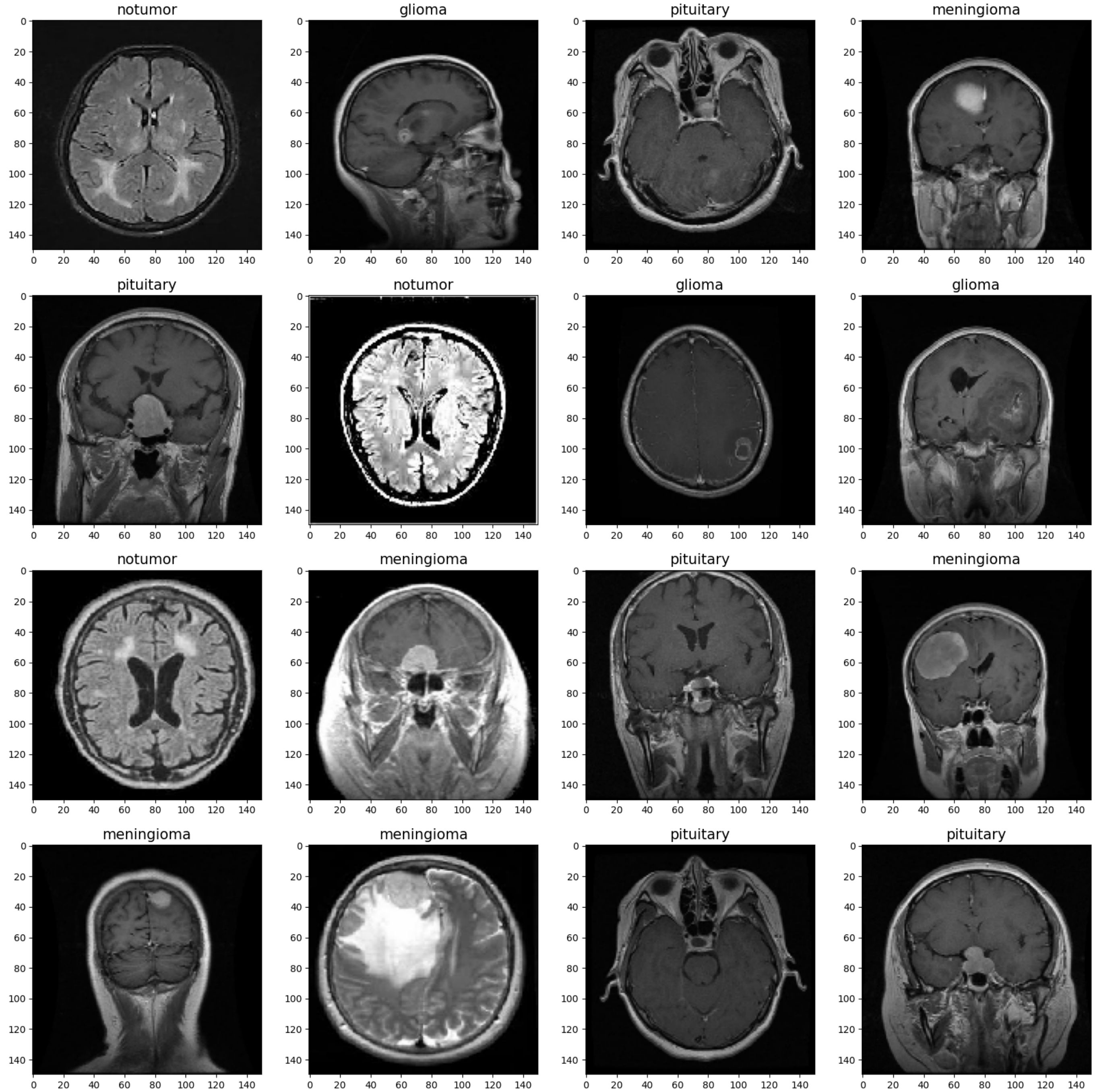
2.3 Preprocesamiento de los datos

```
batch_size = 8
img_size = (150, 150)

_gen = ImageDataGenerator(rescale=1/255,
                          brightness_range=(0.8, 1.2))

ts_gen = ImageDataGenerator(rescale=1/255)

tr_gen = _gen.flow_from_dataframe(tr_df, x_col='Class Path',
                                  y_col='Class', batch_size=batch_size,
                                  target_size=img_size, class_mode='categorical')

valid_gen = _gen.flow_from_dataframe(valid_df, x_col='Class Path',
                                     y_col='Class', batch_size=16,
                                     target_size=img_size, class_mode='categorical')

ts_gen = ts_gen.flow_from_dataframe(ts_df, x_col='Class Path',
                                    y_col='Class', batch_size=16,
                                    target_size=img_size, class_mode='categorical',
                                    shuffle=False)
```

```
Found 400 validated image filenames belonging to 4 classes.
Found 200 validated image filenames belonging to 4 classes.
Found 200 validated image filenames belonging to 4 classes.
```

2.4 Mostramos ejemplos de los datos

```
class_dict = tr_gen.class_indices
classes = list(class_dict.keys())
images, labels = next(ts_gen)

plt.figure(figsize=(20,20))

for i, (image, label) in enumerate(zip(images, labels)):
  plt.subplot(4, 4, i+1)
  plt.imshow(image)
  class_name = classes[np.argmax(label)]
  plt.title(class_name, color='k', fontsize=15)

plt.show()
```



2.5 Construimos el modelo de Deep Learning

```
img_shape=(150, 150, 3)
base_model = tf.keras.applications.Xception(include_top=False, weights='imagenet',
                                             input_shape=img_shape, pooling='max')

model = Sequential([
  base_model,
  Flatten(),
  Dropout(rate=0.3),
  Dense(128, activation='relu'),
  Dropout(rate=0.25),
  Dense(4, activation='softmax')
])

model.compile(Adamax(learning_rate=0.001),
              loss='categorical_crossentropy',
              metrics=['accuracy',
                       Precision(),
```

```
                       Recall()])
model.summary()
```

```
Model: "sequential"
```

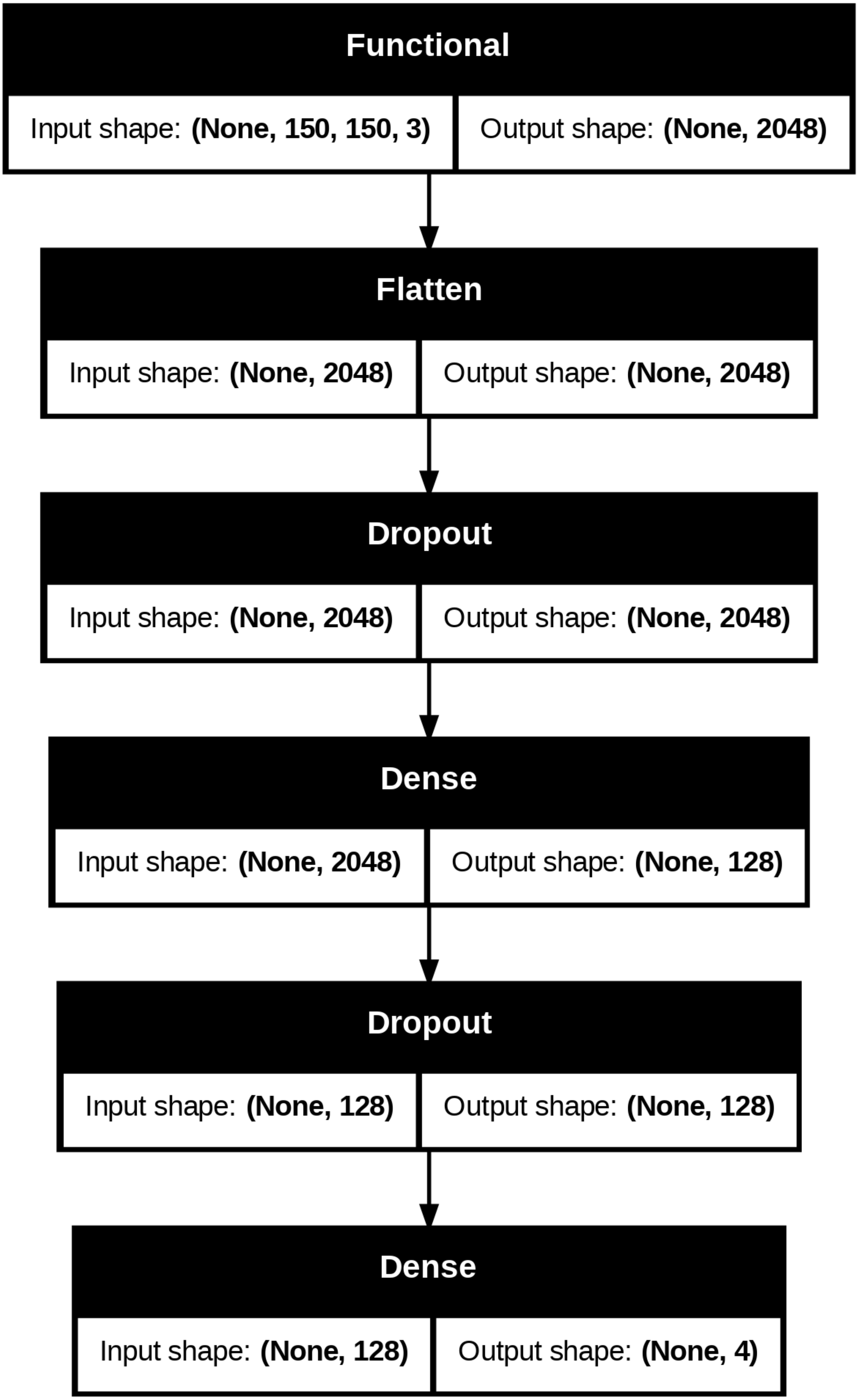
```
 Layer (type)                                  Output Shape                             Param #

 xception (Functional)                         (None, 2048)                          20,861,480

 flatten (Flatten)                             (None, 2048)                                   0

 dropout (Dropout)                             (None, 2048)                                   0

 dense (Dense)                                 (None, 128)                              262,272

 dropout_1 (Dropout)                           (None, 128)                                    0

 dense_1 (Dense)                               (None, 4)                                    516
```

```
 Total params: 21,124,268 (80.58 MB)
```

```
 Trainable params: 21,069,740 (80.37 MB)
```

```
 Non-trainable params: 54,528 (213.00 KB)
```

```
tf.keras.utils.plot_model(model, show_shapes=True)
```


Functional
Input shape: (None, 150, 150, 3)
Output shape: (None, 2048)
Flatten
Input shape: (None, 2048)
Output shape: (None, 2048)
Dropout
Input shape: (None, 2048)
Output shape: (None, 2048)
Dense
Input shape: (None, 2048)
Output shape: (None, 128)
Dropout
Input shape: (None, 128)
Output shape: (None, 128)
Dense
Input shape: (None, 128)
Output shape: (None, 4)

3 Entrenamiento

```
hist = model.fit(tr_gen,
                 epochs=10,
                 validation_data=valid_gen,
                 shuffle=False)

hist.history.keys()
```

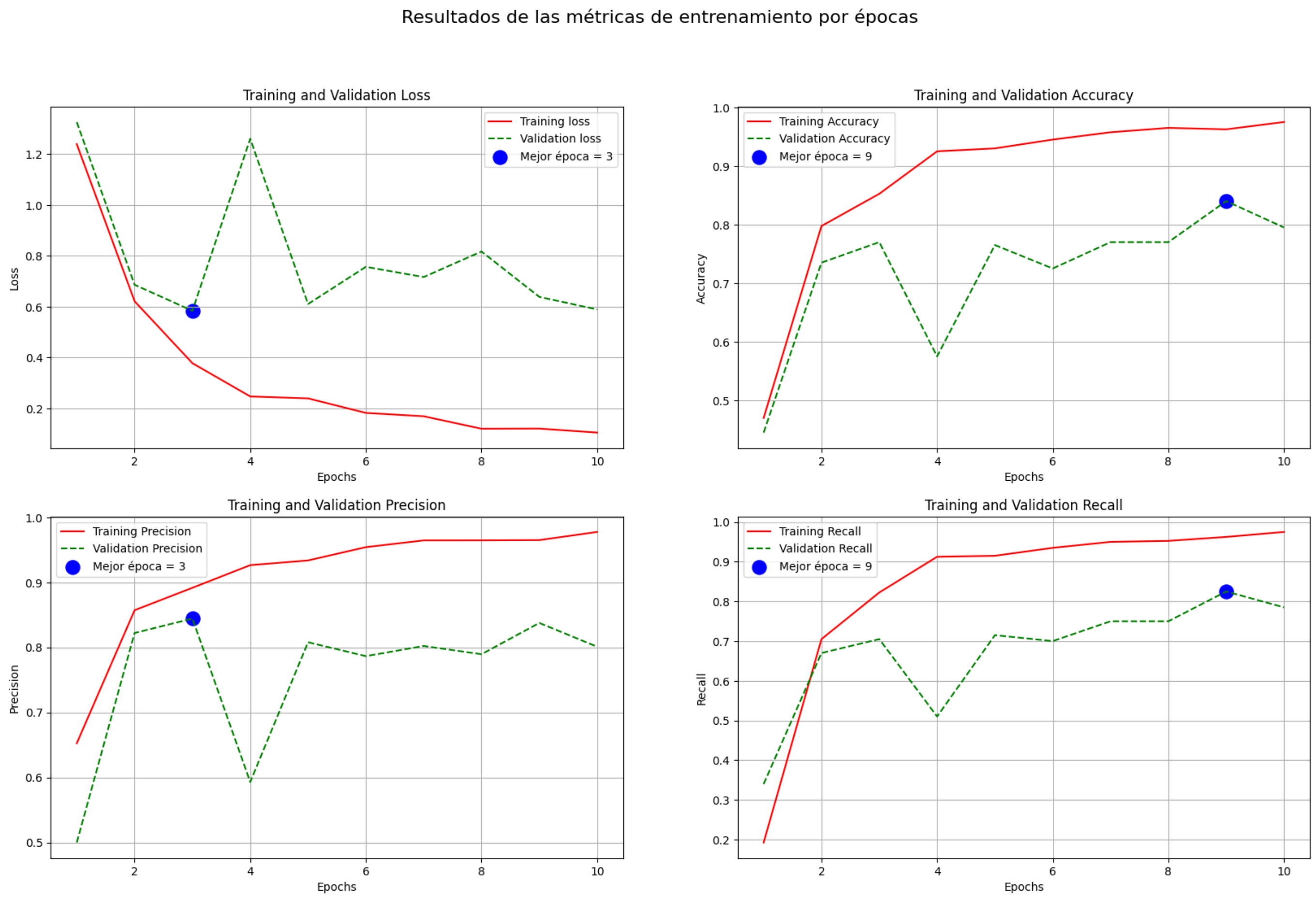
```
Epoch 1/10
50/50            221s 4s/step - accuracy: 0.3521 - loss: 1.6033 - precision: 0.3804 -
   recall: 0.0821 - val_accuracy: 0.4450 - val_loss: 1.3260 - val_precision: 0.5000 -
    val_recall: 0.3400
Epoch 2/10
50/50            223s 4s/step - accuracy: 0.8117 - loss: 0.6516 - precision: 0.8539 -
   recall: 0.6954 - val_accuracy: 0.7350 - val_loss: 0.6861 - val_precision: 0.8221 -
    val_recall: 0.6700
Epoch 3/10
50/50            242s 4s/step - accuracy: 0.8530 - loss: 0.3739 - precision: 0.8901 -
   recall: 0.8227 - val_accuracy: 0.7700 - val_loss: 0.5837 - val_precision: 0.8443 -
    val_recall: 0.7050
Epoch 4/10
50/50            238s 5s/step - accuracy: 0.9245 - loss: 0.2731 - precision: 0.9234 -
   recall: 0.9075 - val_accuracy: 0.5750 - val_loss: 1.2603 - val_precision: 0.5930 -
    val_recall: 0.5100
Epoch 5/10
50/50            154s 3s/step - accuracy: 0.9402 - loss: 0.2144 - precision: 0.9418 -
   recall: 0.9247 - val_accuracy: 0.7650 - val_loss: 0.6102 - val_precision: 0.8079 -
    val_recall: 0.7150
Epoch 6/10
50/50            156s 3s/step - accuracy: 0.9564 - loss: 0.1683 - precision: 0.9631 -
   recall: 0.9483 - val_accuracy: 0.7250 - val_loss: 0.7566 - val_precision: 0.7865 -
    val_recall: 0.7000
Epoch 7/10
50/50            155s 3s/step - accuracy: 0.9546 - loss: 0.1394 - precision: 0.9608 -
   recall: 0.9525 - val_accuracy: 0.7700 - val_loss: 0.7164 - val_precision: 0.8021 -
    val_recall: 0.7500
Epoch 8/10
50/50            155s 3s/step - accuracy: 0.9735 - loss: 0.1229 - precision: 0.9732 -
   recall: 0.9607 - val_accuracy: 0.7700 - val_loss: 0.8169 - val_precision: 0.7895 -
    val_recall: 0.7500
Epoch 9/10
50/50            155s 3s/step - accuracy: 0.9684 - loss: 0.1189 - precision: 0.9705 -
   recall: 0.9684 - val_accuracy: 0.8400 - val_loss: 0.6384 - val_precision: 0.8376 -
    val_recall: 0.8250
Epoch 10/10
50/50            154s 3s/step - accuracy: 0.9725 - loss: 0.0990 - precision: 0.9735 -
   recall: 0.9725 - val_accuracy: 0.7950 - val_loss: 0.5890 - val_precision: 0.8010 -
    val_recall: 0.7850
```

```
dict_keys(['accuracy', 'loss', 'precision', 'recall', 'val_accuracy', 'val_loss', '
   val_precision', 'val_recall'])
```

3.1 Visualizamos el rendimiento del modelo

```
tr_acc = hist.history['accuracy']
tr_loss = hist.history['loss']
tr_per = hist.history['precision']
tr_recall = hist.history['recall']
val_acc = hist.history['val_accuracy']
val_loss = hist.history['val_loss']
val_per = hist.history['val_precision']
val_recall = hist.history['val_recall']
```

```python
index_loss = np.argmin(val_loss)
val_lowest = val_loss[index_loss]
index_acc = np.argmax(val_acc)
acc_highest = val_acc[index_acc]
index_precision = np.argmax(val_per)
per_highest = val_per[index_precision]
index_recall = np.argmax(val_recall)
recall_highest = val_recall[index_recall]

Epochs = [i + 1 for i in range(len(tr_acc))]
loss_label = f"Mejor época = {str(index_loss + 1)}"
acc_label = f"Mejor época = {str(index_acc + 1)}"
per_label = f"Mejor época = {str(index_precision + 1)}"
recall_label = f"Mejor época = {str(index_recall + 1)}"

plt.figure(figsize=(20, 12))

plt.subplot(2, 2, 1)
plt.plot(Epochs, tr_loss, 'r', label='Training loss')
plt.plot(Epochs, val_loss, 'g--', label='Validation loss')
plt.scatter(index_loss + 1, val_lowest, s=150, c='blue', label=loss_label)
plt.title('Training and Validation Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.grid(True)

plt.subplot(2, 2, 2)
plt.plot(Epochs, tr_acc, 'r', label='Training Accuracy')
plt.plot(Epochs, val_acc, 'g--', label='Validation Accuracy')
plt.scatter(index_acc + 1, acc_highest, s=150, c='blue', label=acc_label)
plt.title('Training and Validation Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()
plt.grid(True)

plt.subplot(2, 2, 3)
plt.plot(Epochs, tr_per, 'r', label='Training Precision')
plt.plot(Epochs, val_per, 'g--', label='Validation Precision')
plt.scatter(index_precision + 1, per_highest, s=150, c='blue', label=per_label)
plt.title('Training and Validation Precision')
plt.xlabel('Epochs')
plt.ylabel('Precision')
plt.legend()
plt.grid(True)

plt.subplot(2, 2, 4)
plt.plot(Epochs, tr_recall, 'r', label='Training Recall')
plt.plot(Epochs, val_recall, 'g--', label='Validation Recall')
plt.scatter(index_recall + 1, recall_highest, s=150, c='blue', label=recall_label)
plt.title('Training and Validation Recall')
plt.xlabel('Epochs')
plt.ylabel('Recall')
plt.legend()
plt.grid(True)

plt.suptitle('Resultados de las métricas de entrenamiento por épocas', fontsize=16)
plt.show()
```



4 Testing y Evaluación

4.1 Evaluación

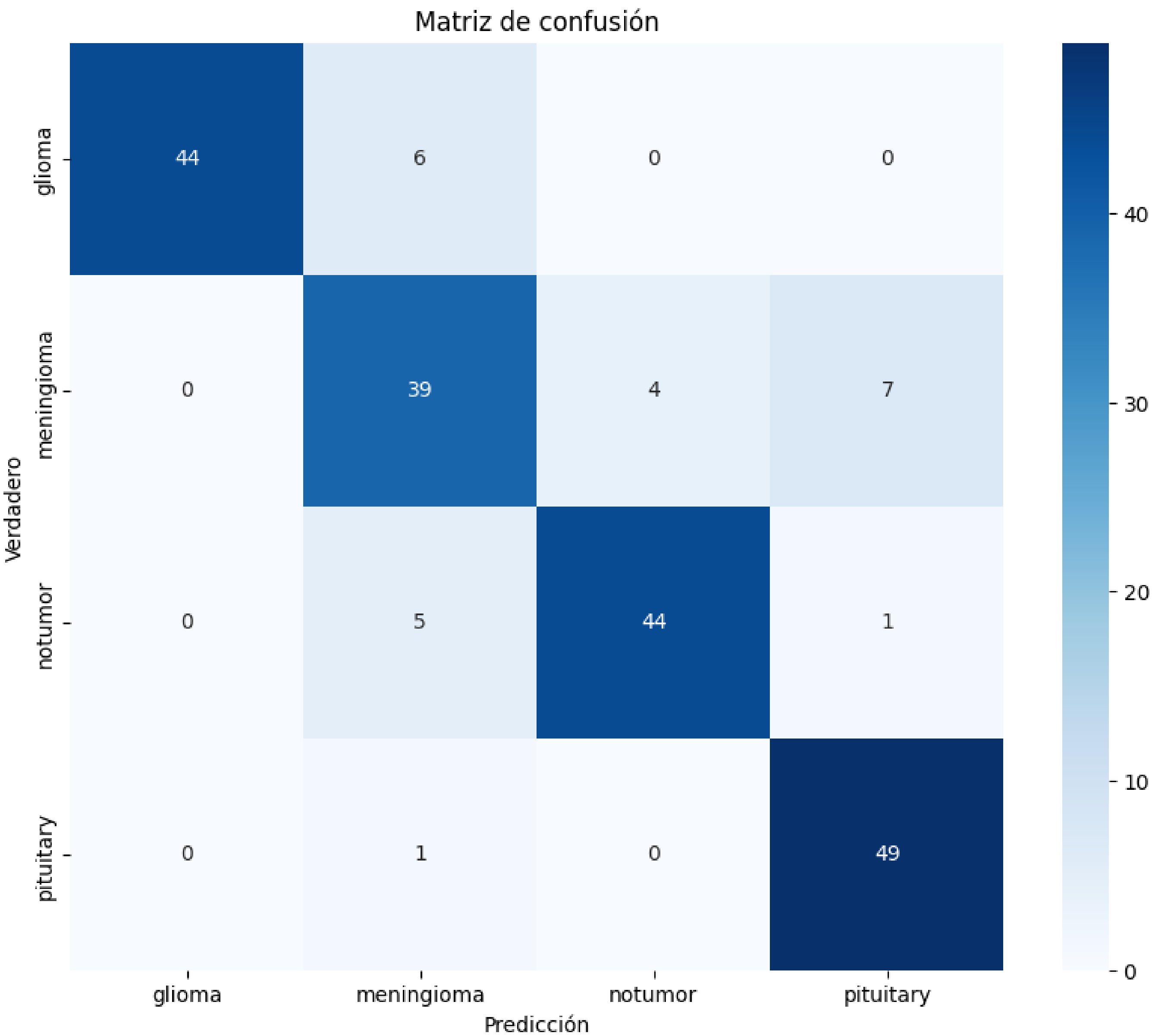
```python
train_score = model.evaluate(tr_gen, verbose=1)
valid_score = model.evaluate(valid_gen, verbose=1)
test_score = model.evaluate(ts_gen, verbose=1)

print(f"Train loss: {train_score[0]:.4f}")
print(f"Train accuracy: {train_score[1]*100:.2f}%")
print('-' * 20)
print(f"Validation loss: {valid_score[0]:.4f}")
print(f"Validation accuracy: {valid_score[1]*100:.2f}%")
print('-' * 20)
print(f"Test loss: {test_score[0]:.4f}")
print(f"Test accuracy: {test_score[1]*100:.2f}%")
```

```
50/50              35s 707ms/step - accuracy: 0.9979 - loss: 0.0221 - precision: 0.9979
    - recall: 0.9892
13/13              18s 1s/step - accuracy: 0.8177 - loss: 0.5431 - precision: 0.8246 -
    recall: 0.8087
13/13              17s 1s/step - accuracy: 0.8652 - loss: 0.4907 - precision: 0.8645 -
    recall: 0.8599
Train loss: 0.0282
Train accuracy: 99.00%
--------------------
Validation loss: 0.5663
Validation accuracy: 81.00%
--------------------
Test loss: 0.4129
Test accuracy: 88.00%
```

```
preds = model.predict(ts_gen)
y_pred = np.argmax(preds, axis=1)
```

```
13/13           18s 1s/step
```

```
cm = confusion_matrix(ts_gen.classes, y_pred)
labels = list(class_dict.keys())
plt.figure(figsize=(10, 8))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=labels, yticklabels=
    labels)
plt.xlabel('Predicción')
plt.ylabel('Verdadero')
plt.title('Matriz de confusión')
plt.show()
```




```
clr = classification_report(ts_gen.classes, y_pred)
print(clr)
```

```
              precision    recall  f1-score   support

           0       1.00      0.88      0.94        50
           1       0.76      0.78      0.77        50
           2       0.92      0.88      0.90        50
           3       0.86      0.98      0.92        50
```

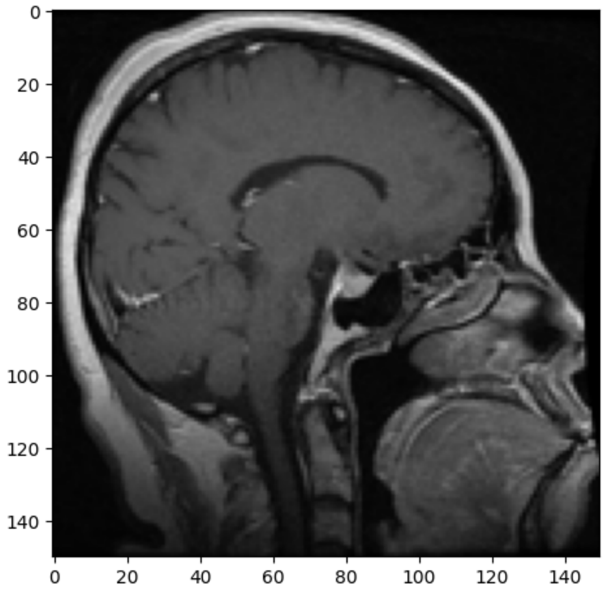
```
    accuracy                           0.88       200
   macro avg       0.89      0.88      0.88       200
weighted avg       0.89      0.88      0.88       200
```

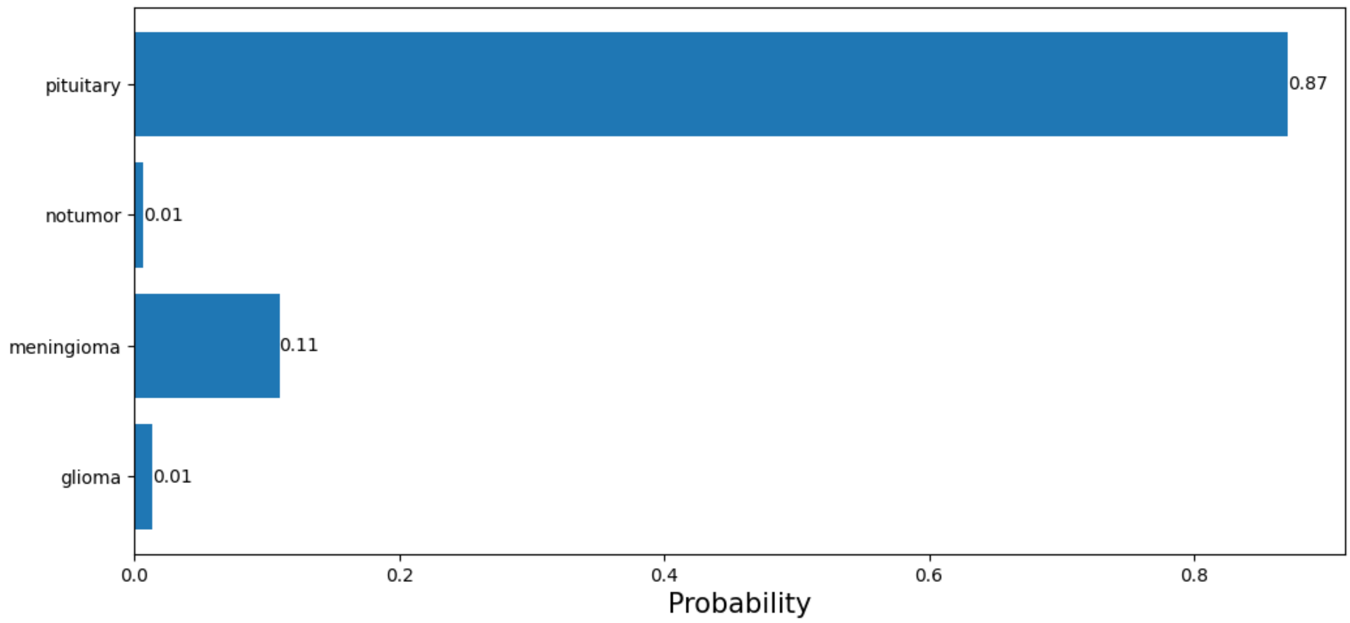
4.2 Testing

```python
def predict(img_path):
    import numpy as np
    import matplotlib.pyplot as plt
    from PIL import Image
    label = list(class_dict.keys())
    plt.figure(figsize=(12, 12))
    img = Image.open(img_path).convert('RGB')
    resized_img = img.resize((150, 150))
    img = np.asarray(resized_img)
    img = np.expand_dims(img, axis=0)
    img = img / 255
    predictions = model.predict(img)
    probs = list(predictions[0])
    labels = label
    plt.subplot(2, 1, 1)
    plt.imshow(resized_img)
    plt.subplot(2, 1, 2)
    bars = plt.barh(labels, probs)
    plt.xlabel('Probability', fontsize=15)
    ax = plt.gca()
    ax.bar_label(bars, fmt = '%.2f')
    plt.show()
```

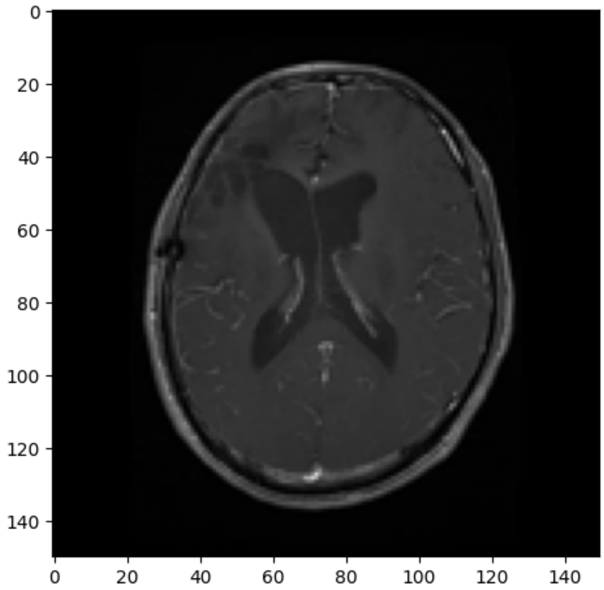
```python
# Directorio base de las imágenes de prueba
test_dir = 'kaggle/input/brain-tumor-mri-dataset/1/Testing'

# Obtener una lista de todas las subcarpetas (clases)
class_folders = [f for f in os.listdir(test_dir) if os.path.isdir(os.path.join(
    test_dir, f))]

# Seleccionar 5 imágenes aleatorias de diferentes clases
random_images = []
for folder in class_folders:
    # Obtener una lista de todas las imágenes en la subcarpeta
    images = [f for f in os.listdir(os.path.join(test_dir, folder)) if os.path.isfile
        (os.path.join(test_dir, folder, f))]
    # Seleccionar una imagen aleatoria de la subcarpeta
    random_image = random.choice(images)
    random_images.append(os.path.join(test_dir, folder, random_image))

# Limitar la lista a 5 imágenes
random_images = random_images[:5]

# Mostrar las imágenes y sus predicciones
for image_path in random_images:
    predict(image_path)
```

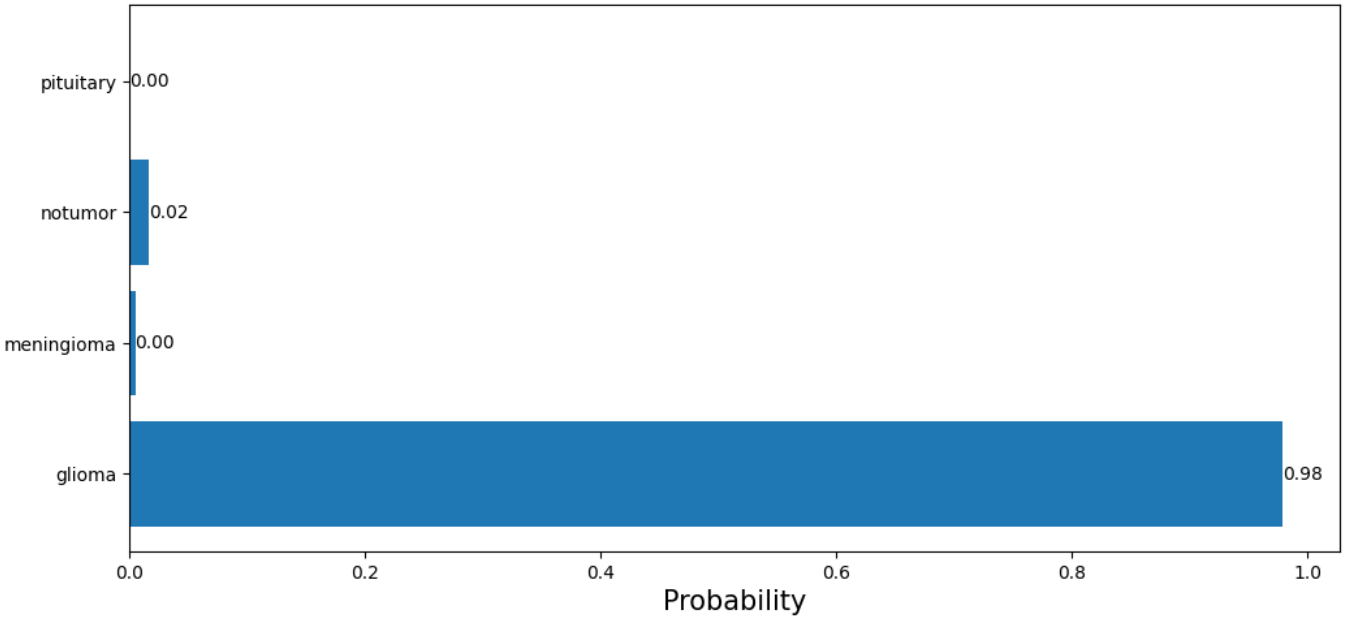
```
1/1            1s 1s/step
```



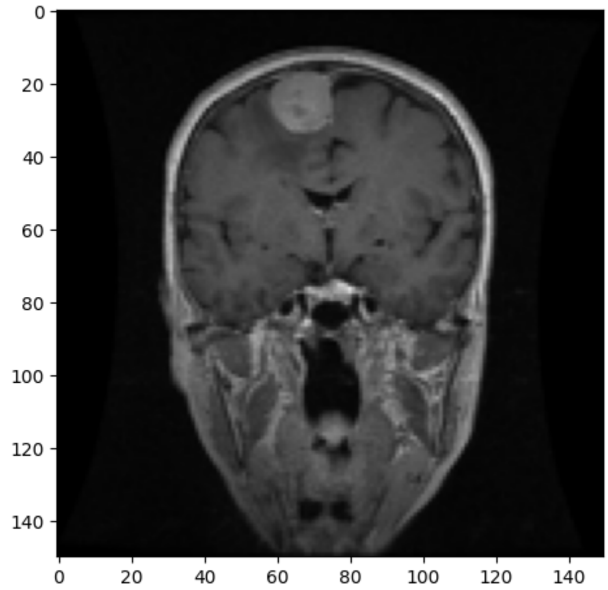


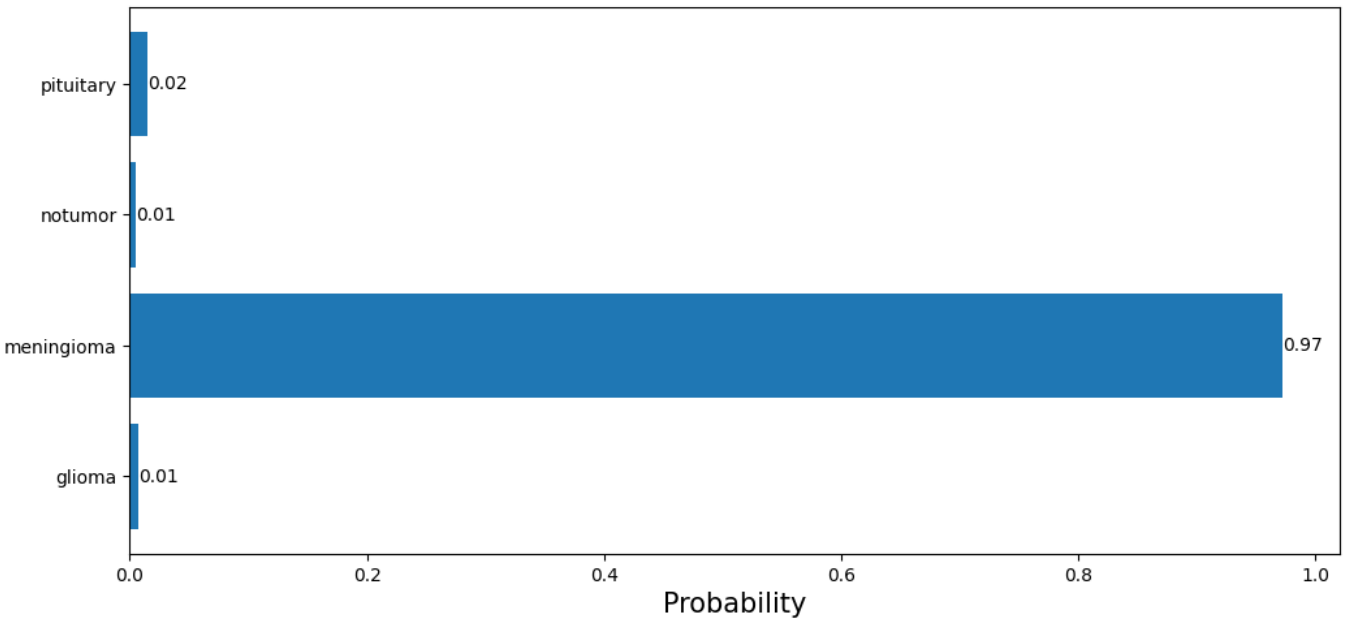

```
1/1              0s 131ms/step
```



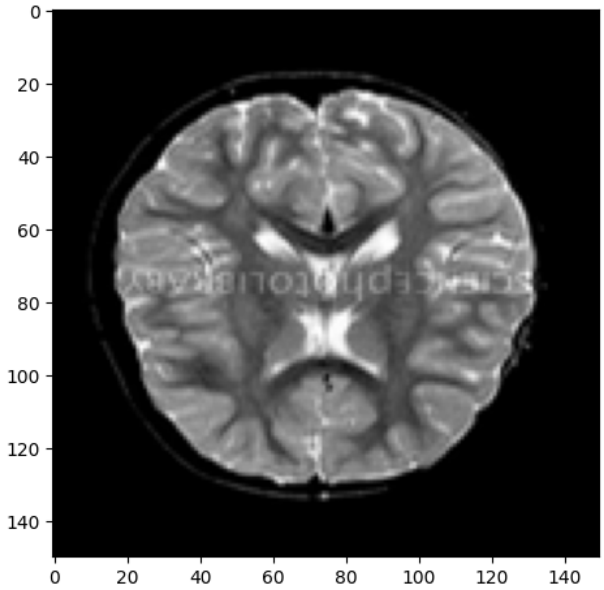


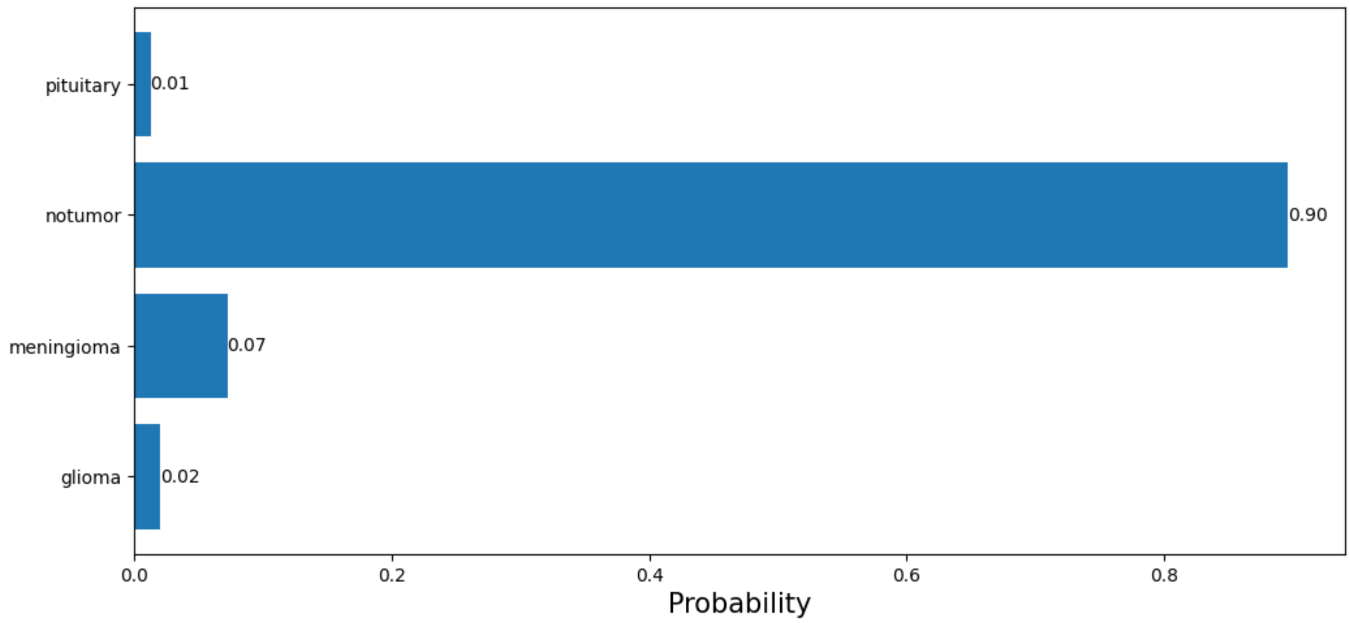

```
1/1            0s 142ms/step
```






```
1/1            0s 124ms/step
```


0
20
40
60
80
100
120
140


pituitary
notumor
meningioma
glioma
0.01
0.90
0.07
0.02
0.0
0.2
0.4
0.6
0.8
Probability