data_preprocessing.ipynb

C: > Users > markn > OneDrive > Documents > Documents > Internships > Artificial Intelligence Internship > Task 2 Computer Vision Project > medical_image_classification > notebooks > data_preprocessing.ipynb > ...

+ Code + Markdown | Run All Clear All Outputs | Outline ··· Select Kernel

Preprocessing

Create a Jupyter notebook (data_preprocessing.ipynb) for preprocessing images and a Python script (data_preprocessing.py) to automate the preprocessing pipeline.

raw

```python
# Imports
import os
import numpy as np
import matplotlib.pyplot as plt
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from src.utils import load_data, plot_sample_images
```

[10] Python

```python
# Paths to data directories
train_dir = 'data/train'
test_dir = 'data/test'
val_dir = 'data/val'

# Load and preprocess data
img_height, img_width = 224, 224
batch_size = 32

train_datagen = ImageDataGenerator(rescale=1./255, rotation_range=20, width_shift_range=0.2,
                                   height_shift_range=0.2, shear_range=0.2, zoom_range=0.2,
                                   horizontal_flip=True, fill_mode='nearest')
test_datagen = ImageDataGenerator(rescale=1./255)
```

[3] Python

```python
train_data = train_datagen.flow_from_directory(train_dir, target_size=(img_height, img_width),
                                               batch_size=batch_size, class_mode='binary')
test_data = test_datagen.flow_from_directory(test_dir, target_size=(img_height, img_width),
                                             batch_size=batch_size, class_mode='binary')
val_data = test_datagen.flow_from_directory(val_dir, target_size=(img_height, img_width),
                                            batch_size=batch_size, class_mode='binary')

# Visualize some sample images
plot_sample_images(train_data)
```

Launchpad ⊗0 ⚠5 0 Connect Go Live tabnine starter See Tabnine Insights Prettier