greatlearning

Project Description – COVID-19 Image classification

Data Description:

You are provided with a training set and a test set of images of 3 classes. Each image has a filename that is its unique id. The dataset comprises 3 classes: COVID-19, Viral Pneumonia, and Normal. The goal of the project is to create a classifier capable of determining the class of the X-ray image.

Dataset:

The project is inspired from a dataset from Kaggle.

Note: For project purposes, the dataset is modified according to the learning requirements.

You are provided with the following files:

- testimage.npy
- testLabels.csv
- trainimage.npy
- trainLabels.csv

The dataset (above 4 files) has to be downloaded from Olympus platform of Great Learning.

Context:

- Can you differentiate an X-ray image of a normal person from an unhealthy one?
- The ability to do so effectively can mean better diagnosis.

Objective:

- To implement the techniques learnt as a part of the course.

Learning Outcomes:

- Pre-processing of image data.
- Visualization of images.
- Building CNN.
- Evaluate the Model.

Steps and tasks:

1. Import the libraries, load dataset, print shape of data, visualize the images in train and test set. (5 Marks) [Hint: To load npy file, use the code: `trainImages = np.load('trainimage.npy')`], It will give you a numpy array. Rest things are as usual (You already how to access numpy arrays.).
2. Explore the data: (5 Marks)
    a. Print the value count of each type of label in:
        i. Training set
        ii. Testing set
    b. Find the number of unique labels in training set and print the names of those.
3. Data Pre-processing: (10 Marks)
    a. Gaussian Blurring. [Hint: Use kernel size of (5, 5)]
    b. Visualize data after performing the Gaussian blurring.
    c. Normalization of data.
4. Make data compatible: (10 Marks)
    a. Reshape data into shapes compatible with Keras models.
    b. Convert labels to one hot vectors.
5. Building CNN: (15 Marks)
    a. Define layers.
    b. Set optimizer and loss function. (Use Adam optimizer and categorical crossentropy.)
6. Fit and evaluate model and print confusion matrix. (10 Marks)
7. Visualize predictions for x_test[2], x_test[3], x_test[33], x_test[36], x_test[59]. (5 Marks)

Happy Learning!