Project Title: Covid-19 Classification Using Computer Vision and Deep Learning [Small Dataset]
Project Team: Karn Jongnarangsin, Yashashvini Rachamallu, Ruchika Gupta

Problem Statement:

Acquiring massive datasets and training the classifier is particularly challenging in the medical profession. The COVID-19 pandemic has brought attention to the necessity of precise and effective diagnostic instruments. Promising outcomes have been observed while identifying COVID-19 from medical photos using Deep Learning algorithms. However, our COVID-19 dataset which consists of about 310 photos that are divided into three classes is what we wish to focus on for this project. Here's where Generative Adversarial Networks (GANs) come into play; they can produce synthetic images that bear resemblance to the original photos while maintaining the integrity of the dataset. After augmenting these synthetic images to increase the size of the dataset, we would like to apply convolutional neural networks (CNNs) with different architectures for the classification task. This project aims to explore deep learning models for COVID-19 classification by implementing convolutional neural networks (CNNs) of varying architectures in PyTorch and training them on a dataset of MRI scans as well as including synthetic images generated by GANs. Crucially, this project is well-suited for the given timeline, as both GANs and CNN models to be implemented are well-established and widely recognized in the field. Furthermore, several pre-trained models are available, which can expedite the development process. This project aims to provide a practical solution for COVID-19 image classification, even when confronted with a limited dataset, contributing to more accurate and timely diagnoses in the medical profession.

Objectives:

1. Implement GANS (DCGAN) to generate synthetic images from the dataset.
2. Implement COVID-19 classification models that can accurately classify COVID-19 in MRI scans.
3. Implement preprocessing, training, and evaluation pipelines in PyTorch.
4. Assess the individual model performances using relevant evaluation metrics.
5. Compare the models based on their performance metrics as well as computational cost/complexity.

Methodology:

1. **Data Collection:** Obtain a dataset from Kaggle, a well-known collection of datasets. The specific dataset that we are using consists of MRI scan images of lungs with corresponding labels.

2. **Data Preprocessing:** Normalize, resize the images, followed by the image augmentation in two methods. One method uses the traditional techniques like image rotation, image flipping etc. Whereas the other method uses GAN's to generate synthetic images.
3. **Model Architecture:** Implementing the GAN's for generating the synthetic images, utilizing DCGAN architecture, followed by implementation of convolutional neural networks (CNNs) for image classification, utilizing the VGG, ResNet, and EfficientNet architectures.

4. **Loss Function:** Employ cross-entropy loss as the optimization objective.

5. **Training and Validation:** Create a training and testing data split. Train the model on the training set while monitoring its performance on a validation set. Apply regularization techniques to prevent overfitting.

6. **Post-processing:** Implement post-processing steps to refine models' outputs.

7. **Hyperparameter Tuning:** Tune hyperparameters to potentially improve models' performances.

8. **Evaluation:** Assess models' performances on the test set and compare models in terms of performance and computational cost.

Task Distribution:

1. **Karn Jongnarangsin:**
    a. CNN implementation
    b. Evaluation

2. **Yashashvini Rachamallu:**
    a. GAN implementation
    b. Evaluation

3. **Ruchika Gupta:**
    a. CNN implementation
    b. Evaluation