**Lung and Colon Cancer Classification using CNNs**

**Overview**

This Jupyter Notebook implements a deep learning approach to classify histopathological images of lung and colon cancer using Convolutional Neural Networks (CNNs). The study leverages transfer learning and ensemble learning to enhance classification accuracy.

**Objectives**

* Improve classification accuracy of histopathological images for lung and colon cancer.
* Utilize transfer learning with pre-trained models (VGG-16, ResNet-50, and DenseNet-121).
* Combine multiple models through ensemble learning to enhance diagnostic reliability.

**Dataset**

* **Source**: [Kaggle Dataset - Lung and Colon Cancer Histopathological Images](https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images).
* **Path Used**: "C:/Users/samulya/Desktop/lung\_colon\_image\_set" (Should be replaced with a dynamic dataset path).
* **Preprocessing Steps**:
  + Images are resized to **224x224** pixels for compatibility with pre-trained models.
  + Dataset split: **80% training, 10% validation, 10% testing**.
  + Images normalized before training.

**Methodologies**

**1. Transfer Learning (CNN)**

* Pre-trained CNN models (**VGG-16, ResNet-50, DenseNet-121**) are used to extract features.
* Fine-tuning is performed to adapt models to the cancer classification task.

**2. Ensemble Learning**

* Predictions from all three models are averaged to form a robust ensemble model.
* This method helps in reducing model-specific biases and improving accuracy.

**Implementation Steps**

**1. Import Required Libraries**

* TensorFlow and Keras for deep learning.
* ImageDataGenerator for image augmentation.
* Matplotlib and scikit-learn for visualization and evaluation.

**2. Data Loading and Preprocessing**

* Images are loaded using tf.keras.utils.image\_dataset\_from\_directory().
* Labels are one-hot encoded (label\_mode='categorical').
* Data augmentation is applied to improve generalization.

**3. Model Architecture**

* Uses three pre-trained models (VGG-16, ResNet-50, DenseNet-121).
* Averages their predictions to generate ensemble outputs.
* Optimizer: **Adam**
* Loss Function: **Categorical Crossentropy**
* Callback: **EarlyStopping** to prevent overfitting.

**4. Model Evaluation**

* Performance is measured using:
  + **Confusion Matrix**
  + **Classification Report**
  + **Accuracy and Loss Curves**

**Significance**

* The ensemble model provides a **more reliable diagnosis** compared to individual CNNs.
* Automates cancer detection to **assist medical professionals in early diagnosis**, improving survival rates.