Lung and Colon Cancer Classification Using Custom CNN.

* Dataset: <https://github.com/tampapath/lung_colon_image_set>
  1. **Dataset description: -** The dataset contains 25,000 color images with five classes of 5,000 images each. All images are 768 x 768 pixels in size and are in jpeg file format. Our dataset can be downloaded as a 1.85 GB zip file LC25000.zip. After unzipping, the main folder lung\_colon\_image\_set contains two subfolders: colon\_image\_sets and lung\_image\_sets. The subfolder colon\_image\_sets contain two secondary subfolders: colon\_aca subfolder with 5,000 images of colon adenocarcinomas and colon\_n subfolder with 5,000 images of benign colonic tissues. The subfolder lung\_image\_sets contain three secondary subfolders: lung\_aca subfolder with 5,000 images of lung adenocarcinomas, lung\_scc subfolder with 5,000 images of lung squamous cell carcinomas, and lung\_n subfolder with 5,000 images of benign lung tissues.
  2. **Image acquisition: -** HIPAA compliant and validated seven hundred fifty total images of lung tissue (250 benign lung tissue, 250 lung adenocarcinomas, and 250 lung squamous cell carcinomas) and 500 total images of colon tissue (250 benign colon tissue and 250 colon adenocarcinomas) Borkowski et al 2 were captured from pathology glass slides as we previously described.
  3. **Image augmentation: -** All images were cropped to square sizes of 768 x 768 pixels from original 1024 x 768 pixels using python programming language. Subsequently, images were augmented using the Augmentor software package. Augmentor is an image augmentation library in Python for machine learning. It aims to be a standalone library that is platform and framework independent, which is more convenient, allows for finer-grained control over augmentation, and implements the most real-world relevant augmentation techniques. It employs a stochastic approach using building blocks that allow for operations to be pieced together in a pipeline.[9] Using Augmentor, we expanded our dataset to 25,000 images by the following augmentations: left and right rotations (up to 25 degrees, 1.0 probability) and by horizontal and vertical flips (0.5 probability).

Classification Workflow:

# 1, Import Libraries

# 2, Loading Dataset

# 3, Split dataset

# 4, Gettting images into tensors

# 5, Creating dataloaders

# 6, Building the model

# 7, **Setting the cost function and the optimizer**

# 8, Training the model

# 9, **Model Evaluation**

# **10, Show predictions**

* There are five classes in the dataset, each with 5,000 images, being:
* Lung benign tissue
* Lung adenocarcinoma
* Lung squamous cell carcinoma
* Colon adenocarcinoma
* Colon benign tissue