Lung and Colon Cancer Detection

Team members:

- Mohamed Abdalkader
- Mohammed Ahmed Ismail
- Mohammed Ahmed Mahmoud
- Ahmed Mohammed Elsayed
- Ahmed Said Emam
- Ahmed Maher

Introduction:

This graduation project focuses on developing a computer vision system to classify histopathological images of colon and lung tissue. The main goal is to utilize deep learning techniques to accurately classify images into specific classes, thus helping in diagnosing lung and colon cancerous tissues faster, which helps in early detection of tumors and contributes to initiating early treatment for the patient.

Objectives:

- Develop a robust computer vision system for classifying colon and lung histopathological images.
- Fine-tune a DenseNet model using training data.
- Evaluate the system's performance using accuracy, precision, recall, and F1 score metrics.
- Present the project findings and recommendations for future work.

Scope:

- Data Collection and Preprocessing: Gather and preprocess histopathological images for analysis.
- Model Development: Build the CNN and DenseNet models using TensorFlow or Keras.
- Performance Validation: Utilize a confusion matrix and other relevant metrics to validate performance.
- Model Deployment and experiment: Deploy the model using flask API and run experiments using MLflow.

Project Plan:

Week 1: Data Collection and Preprocessing

- **Tasks:**
  - ***Data collection***:
    - A dataset from Kaggle was used. It is a collection of histopathological images of colon and lung consisting of 5 classes, each class consisting of 250 images. Augmenter package was used to increase the number of images to 5000 images per class (i.e. total images are 25000 images).
  - ***Data Preprocessing***:
    - Data was cleaned and pre-processed, images were resized to 200*200, RGB channel was used once, and gray channel was used once before being used in the model.
  - ***Tools***:
    - Python [Pandas, NumPy, cv2, matplotlib, train_test_split, ImageData-Generator].

- **Deliverables*:***
  - Cleaned and preprocessed dataset's Shapes.
  - Dataset was splitted into Train: 60%, validation: 20%, Test: 20%.
  - Data preprocessing notebook.

Week 2: Machine Learning Model

- **Tasks:**
  - ***CNN Model:***
    1. Building a mode to identify and classify colon and lung tumors using deep learning model.
    2. Training the model on dataset images (grayscale and colored) and evaluating the mode.
    3. Developing the model and improving its evaluation.
  - ***Tools:***
    a. Python (tensorflow).

- **Deliverables*:***
  - Colon and Lung Cancer Detection models and performance metrics.

Week 3: Advanced Techniques

- **Tasks:**
  - *DenseNet model:*
    - Developed two DenseNet models, one model was trained on grayscale images, the other model utilized colored images with transfer learning from a pre-trained DesenNet.
    - Using DenseNet and training a model with both grayscale and color images, we comprehensively analyzed the effect of image color on cancer detection performance.
    - After evaluating both models, the grayscale DenseNet model emerged as the best performer, achieving 99% accuracy on the test dataset.
  - *Tools:*
    - Python (tensorflow/DenseNet).

- **Deliverables:**
  - Enhanced colon and lung cancer detection model with DenseNet.

Week 4: MLOps, Deployment

- **Tasks:**
  - **MLOps**: Use MLflow to manage and track the model training and experiments for classifying histopathological images.
  - **Flask Deployment**: Develop a Flask application to deploy the trained model and provide a user-friendly interface for image classification.
  - **Final Report and Presentation**: Document the project, including model development, deployment, and the results of MLflow experiments.
  - **Tools**:
    - MLflow.
    - Python (TensorFlow/keras).
    - Flask for deployment.