Amir Aynede
Department of Pharmacy and Biotechnology (FaBiT)
Alma Mater Studiorum – University of Bologna

Project Proposal: Deep Learning for Cancer Histopathology Image Classification using LC25000

Cancer remains one of the most significant global health challenges, and accurate diagnosis plays a critical role in effective treatment and improved patient outcomes. In clinical practice, histopathology—the microscopic examination of tissue samples—is the gold standard for cancer diagnosis. However, manual analysis of histopathology slides is time-consuming, labor-intensive, and subject to inter-observer variability. In recent years, deep learning, particularly convolutional neural networks (CNNs), has shown great promise in automating and enhancing diagnostic accuracy in medical imaging.

For this project, I propose developing an end-to-end deep learning pipeline for cancer subtype classification using the publicly available **LC25000 dataset**. This dataset comprises 25,000 histopathology images evenly distributed across five classes: Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Benign Lung Tissue, Colon Adenocarcinoma, and Benign Colon Tissue. The images are preprocessed to a standardized resolution (224×224 pixels), making the dataset well-suited for transfer learning applications.

The objective of this project is to build a robust CNN-based classifier capable of distinguishing between these cancer subtypes and benign tissues. I will use a pretrained **ResNet18** model as the base architecture and fine-tune it for our classification task. The model will be trained and evaluated using performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix. To ensure interpretability, I will incorporate **Grad-CAM** (Gradient-weighted Class Activation Mapping) visualizations to highlight the image regions most influential in the model's decision-making process.

The project will follow a modular code structure consisting of separate scripts for data loading, model definition, training, evaluation, and visualization. Additionally, a Jupyter notebook will be included for experimentation and demonstration of results. The implementation will be done in Python using PyTorch, and the entire codebase will be well-documented and reproducible. This project aims to contribute to the development of intelligent diagnostic tools that can aid pathologists and improve clinical workflows. Upon your approval, I will begin implementing the pipeline and conducting experiments using the LC25000 dataset.