Assignment 3 (20%)

Problem

Pneumonia is an acute respiratory infection caused by viruses, bacteria, or fungi that results in the air sacs filling with pus or fluid. Symptoms include a cough with mucus, fever, chills, and breathing difficulties. If left untreated, pneumonia can be life-threatening, particularly for high-risk groups such as infants, young children, and the elderly. In 2019, nearly 2.5 million people worldwide died from pneumonia, as reported by Our World in Data (https://ourworldindata.org/pneumonia). In the United States, common causes of viral pneumonia include influenza viruses, respiratory syncytial virus (RSV), and SARS-CoV-2 (COVID-19).

Diagnosing pneumonia often involves chest X-rays, alongside clinical exams and patient medical histories. Advances in AI have shown promise in detecting abnormal patterns on chest X-rays, addressing the need for more specialists to review radiographs.

This problem explores a visual anomaly detection example using a chest X-ray image dataset, available at https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. You will learn about using a pre-trained VGG16 model for visual anomaly detection, then employ the XAI method to highlight abnormal areas in the images and explain model predictions.

The chest X-ray dataset includes 5,864 images divided into three main folders (train, test, val) and two subfolders (PNEUMONIA, NORMAL). These images, collected from routine clinical care, were graded by expert physicians for use in AI applications.

Objectives

1. As a data scientist, you are assigned the task of building an ML model to determine if PNEUMONIA is present or not i.e., if the target is 1 or 0.
2. As a data scientist, you want to understand:
    a. What is the overall logic of the model in making decisions?
    b. Is the logic reasonable, so that the hospital can deploy the model with confidence?

About the dataset:

Using a chest X-ray image dataset, available at https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

Using the Colab template given or develop your own codes, complete the following tasks**:**

Task 1: (5 marks)

- Import Libraries
- Load the images
- Separate the Input and Target Features of the images
- Split the data into Train and Test Sets
- Scale your data if required

Task 2: (5 marks)

- Train and compare three Machine Learning/Deep Learning Models
- Evaluate the model using F1 Score

Task 3: (5 marks)

- Select and use a suitable XAI method (GradCAM, RISE or LIME) to explain the Deep Learning or Machine Learning model.

Task 4: (5 marks)

- Write your XAI explanation.

Task 5: (5 marks)

- Select and use another XAI method (different from Task 3) to explain the Deep Learning or Machine Learning model.

Task 6: (5 marks)

- Write your XAI explanation for Task 5. Visual outputs are required (heatmaps, feature importance, textual explanations, etc.).