Enhancing Pneumonia Detection in X-Rays with Deep Learning Project Overview

Due Date: TBD
Submission Format: Upload a PDF and a link to your GitHub repository on Canvas.

Overview:
You will submit a professional case study consisting of two parts:

- A **written report** in PDF format (including a references page)
- A **GitHub repository** with your project code and any relevant data

Purpose: This case study is your opportunity to demonstrate both technical and conceptual data science skills through an integrated, independently-driven project. It mirrors the kind of work you might encounter if you were a data scientist working in a medical facility.

Your Task: Leverage the tools and knowledge you've gained throughout the course to design a model that predicts pneumonia cases in a set of X-Ray images.

Success Criteria: You will be evaluated based on how well you meet the expectations outlined in the rubric below.

Category	**Details**
1. Formatting	1. **Written Report**: Submit a polished PDF document with organized sections and visualizations. 2. **GitHub Repository**: ● Include all code files and any necessary data (excluding large raw datasets). ● Organize your folders (e.g., `/models`, `/notebooks`, `/data`, `/scripts`) and include a `README.md` with setup instructions. ● Title your repository: `CS-PneumoniaDetection-[First NameLastName]` 3. **References**:

	● Include citations on a separate reference page in the PDF. ● Use IEEE citation style for any sources not already provided
2. Written Report	Your written analysis should communicate your approach, methods, and conclusions. ● **Problem Definition**: ○ Summarize the clinical relevance of pneumonia detection and the importance of early diagnosis. ● **Approach Overview**: ○ Describe your strategy for preprocessing DICOM files, resizing, normalization, and how you adapted them for CNN input. ○ Include a diagram or flowchart showing your end-to-end workflow. ● **Model Architecture**: ○ Outline your CNN structure, training settings, and hyperparameters. ● **Results & Evaluation**: ○ Report accuracy, loss curves, confusion matrix, or other relevant metrics. ○ Discuss any observations about overfitting, model limitations, or misclassifications. ● **Reflection**: ○ Reflect on challenges encountered (e.g., working with DICOM files, model tuning).

	○ Describe what you'd improve or change and lessons learned from the project.
3. Code	Your analysis should include: ● **Data Preprocessing**: ○ Code for loading and converting DICOM files ○ Image resizing (e.g., to 224x224) and normalization to [0, 1] ○ Channel expansion for grayscale images ● **Model Development**: ○ CNN implementation with clear comments ○ Training loop, validation tracking, and testing on unseen data ● **Model Evaluation**: ○ Metric calculations (e.g., accuracy, precision, recall) ○ Plots showing training/validation loss and accuracy
4. References	● At the end of your PDF report, include all additional sources (e.g., research papers on CNNs, medical imaging tutorials) in **IEEE citation style**.