Project Proposal: Deep Learning for Automated Disease Detection in Medical Images

Learning Objective: Finalize one of your capstone ideas based on the feedback and discussions with your mentor. Write a proposal that identifies a real-world problem and an approach to solve it.

Problem Statement: Early and accurate detection of diseases such as cancer is crucial for effective treatment and patient outcomes. Despite advances in medical imaging technology, the interpretation of these images remains a time-consuming and error-prone process. Radiologists often face high workloads, which can lead to diagnostic errors and delayed treatment. Automated systems using deep learning can significantly aid radiologists by highlighting areas of concern in medical images, thereby improving diagnostic accuracy and efficiency.

Dataset: To address this problem, we will utilize the following datasets:
- RSNA Pneumonia Detection Challenge dataset (Kaggle): This dataset contains a large set of chest X-ray images with corresponding annotations for pneumonia cases. It provides a robust foundation for training and testing a deep learning model.
- NIH Chest X-ray Dataset: This dataset includes over 100,000 X-ray images with 14 labeled diseases, offering a diverse set of examples for model training and evaluation.

Approach: The project aims to develop a convolutional neural network (CNN) model to detect and classify diseases from chest X-ray images. The specific steps in the process will include:

1. Data Preprocessing:
   - Load and preprocess the chest X-ray images.
   - Normalize the image data.
   - Annotate the images with the appropriate disease labels from the datasets.

2. Model Development:
   - Design and implement a CNN architecture suitable for medical image classification.
   - Train the model using the preprocessed datasets.
   - Employ data augmentation techniques to improve model generalization.

3. Model Evaluation:
   - Evaluate the model's performance using standard metrics such as accuracy, precision, recall, and F1-score.
   - Perform cross-validation to ensure the model's robustness and reliability.

4. Deployment:
   - Develop an API to deploy the trained model for real-time use in clinical settings.
   - Implement a user-friendly interface for radiologists to interact with the system.

5. Validation:

- Conduct a pilot study with radiologists to validate the system's effectiveness in a real-world clinical environment.
- Collect feedback and iteratively improve the model and deployment process.

Presentation: The proposal will be clear and follow a logical flow, ensuring that each section builds on the previous one. The final submission will be a PDF document submitted to the associated Github repo.