Challenge 2

Soil vs Non-Soil Classification

Team Member: Aman Sagar

Approach to Solving the Problem

To build a robust pipeline for soil type classification, I first needed to ensure that the input images were actually soil-related. Any irrelevant or noisy input—such as roads, plants, wooden textures, or man-made materials—would only confuse the soil classification model. Therefore, I designed a binary classification model to identify whether an image contained soil or not.

I chose a ResNet-18 model pre-trained on ImageNet for this task due to its excellent balance between performance and computational efficiency. The dataset was constructed by combining soil images from the main challenge dataset and a manually collected set of non-soil images. These included concrete, plants, and other natural surfaces.

The model was trained using standard PyTorch tools with data augmentation strategies to avoid overfitting and encourage generalization.

Challenges Faced

1. **Ambiguity in non-soil samples**: Some non-soil textures like dry wood or brown rocks looked very similar to soil.

2. **Lack of labeled non-soil data**: The challenge dataset only provided soil types; I had to construct the non-soil dataset from scratch.

3. **Model confusion due to visual similarities**: Colors and textures often caused the model to misclassify.

How I Overcame These Challenges

- I collected and curated a separate set of images using web scraping and public datasets.

- Data augmentation was applied aggressively—rotation, color jitter, flipping—to help the model learn distinguishing features.

- A validation loop with manual error analysis was introduced. Misclassified non-soil images were added back into the training set as hard examples.

- I applied transfer learning with frozen lower layers and gradually unfroze them in later epochs for fine-tuning.

Final Observation and Results

This model served as a pre-filter for the main soil classification system. Though not evaluated on the leaderboard, its role was essential in ensuring only clean, soil-relevant data reached the downstream classifier. The final model achieved **1.0000 accuracy** on the leaderboard.