Detecting Cassava Leaf Diseases Using CNN-Based Image Classification

Introduction:

Cassava is a major food crop in many developing countries, but its production is often threatened by various leaf diseases. Early and accurate disease detection is essential to prevent crop loss and improve food security. Traditional disease identification relies on manual inspection, which is time-consuming and prone to human error.

With the rise of deep learning, **Convolutional Neural Networks (CNNs)** can automatically classify plant leaf images with high accuracy. This project will use deep learning to classify cassava leaf images into five categories, helping to identify diseases early and support farmers in crop management.

Dataset Overview:

The dataset for this project comes from the **Cassava Leaf Disease Classification** Kaggle competition.

- **Total Images:** ~21,000
- **Classes (5):**
    1. Cassava Bacterial Blight (CBB)
    2. Cassava Brown Streak Disease (CBSD)
    3. Cassava Green Mottle (CGM)
    4. Cassava Mosaic Disease (CMD)
    5. Healthy Leaf

All images are RGB and vary in resolution. The dataset presents a **fine-grained classification challenge**, as several diseases have similar visual patterns.

Objective:

The main goal of this project is to **classify cassava leaves into healthy or diseased categories** using a deep learning model. Specific objectives include:

- Load and preprocess the cassava leaf dataset.
- Build and train a CNN-based model (baseline and transfer learning).
- Apply **data augmentation** to improve generalization.
- Evaluate the model using **accuracy, precision, recall, and F1-score**.
- Visualize predictions and misclassifications to understand model behavior.

Technologies and Tools:

- **Language:** Python
- **Libraries:** PyTorch, torchvision, NumPy, pandas, matplotlib, scikit-learn
- **Modeling:** Custom CNN and **Transfer Learning** with ResNet50
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-Score, Confusion Matrix
- **Visualization Tools:** matplotlib, seaborn

Methodology:

1. **Data Preparation**
    - Load dataset and explore sample images for each class.
    - Resize images to 224×224 and normalize pixel values.
    - Split dataset into **training (70%)**, **validation (15%)**, and **testing (15%)**.
    - Apply **data augmentation**: horizontal/vertical flips, rotations, color jitter, and random crops.
2. **Model Building**
    - **Baseline:** Small custom CNN to understand dataset characteristics.
    - **Transfer Learning:** Use **ResNet50 pretrained on ImageNet**, fine-tune the last layers for 5-class classification.
3. **Training the Model**
    - Use **Adam optimizer** and **CrossEntropyLoss**.
    - Train for 20–30 epochs with **early stopping**.
    - Monitor training and validation accuracy and loss.
4. **Evaluation and Testing**
    - Evaluate the trained model on the test set.
    - Generate **classification report** and **confusion matrix**.
    - Visualize **sample predictions** and analyze misclassifications.

Conclusion:

This project will demonstrate how deep learning can be applied to **real-world agricultural problems**. By automatically detecting cassava leaf diseases, this solution could assist farmers in early disease management and improve crop yield. The project also serves as a strong demonstration of CNNs and transfer learning in **fine-grained image classification**.