2025 Term Paper Project HW-1

Kuan-Chen, Chen

Problem Description:

Plant diseases significantly impact agricultural productivity, with traditional manual inspection methods being time-consuming and subjective (Barbedo, 2019). This research develops an automated tomato leaf disease detection system using computational analysis of natural surface structures on plant leaves. The study analyzes leaf surface patterns, texture variations, and visual symptoms to accurately classify healthy and diseased tomato leaves using machine learning techniques (Mohanty et al., 2016).

Scope Definition:

This study focuses on tomato leaf disease detection using three integrated machine learning approaches applied to natural surface structure analysis. The research utilizes the "Tomato leaf disease detection" dataset from Kaggle, derived from the PlantVillage repository (Hughes & Salathé, 2015). The dataset contains ten categories: Tomato mosaic virus, Target Spot, Bacterial spot, Tomato Yellow Leaf Curl Virus, Late blight, Leaf Mold, Early blight, Spider mites, Tomato healthy, and Septoria leaf spot, with 1,000 training images and 100 test images per category. The scope includes: (1) CNN classification of these ten conditions, (2) Regression analysis to predict disease severity scores based on green-to-brown pixel ratios, and (3) Clustering algorithms to group images by severity levels.

Importance and Significance:

Automated plant disease detection systems can revolutionize agricultural practices by enabling early disease identification and reducing crop losses (Singh & Misra, 2017). This research contributes to precision agriculture by developing cost-effective solutions for disease monitoring. The comparative analysis between CNN and traditional machine learning methods will provide insights for selecting optimal approaches based on computational resources and accuracy requirements.

Research Questions:

1. How effectively can CNN classify ten categories of tomato leaf conditions based on natural surface structure patterns?

2. Can regression models accurately predict disease severity scores derived from pixel-based color analysis and disease scaling factors?

3. What distinct patterns emerge when clustering tomato leaf images based on predicted severity scores?

4. How do clustering results relate to CNN classification outcomes for disease pattern analysis?

References:

Barbedo, J. G. A. (2019). Plant disease identification from individual lesions and spots using deep learning. Biosystems Engineering, 180, 96-107.

Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. Information Processing in Agriculture, 4(1), 41-49.