2025_"ShuWei Cup"

Problem B：Research on Crop Leaf Disease

(I) Background

As global agriculture advances into the era of Smart Agriculture 4.0, digital, intelligent, and automated technologies are profoundly reshaping traditional agricultural production models, crop management processes, and agricultural value creation methods. Amid this transformation, modern agricultural equipment such as drone remote sensing, IoT sensors, and intelligent irrigation systems has been widely applied in farmland monitoring, significantly reducing manual inspection costs and improving agricultural management efficiency. However, the true upgrading of agricultural intelligence lies not only in automation but also in predictability—realizing real-time monitoring of crop health, early warning of diseases, and accurate diagnosis through computer vision and AI algorithms, thereby dynamically optimizing plant protection schemes and minimizing pesticide use and yield losses.

Traditional agriculture has long faced pain points such as yield reduction, quality decline, and economic losses caused by crop disease outbreaks. The era of smart agriculture demands a leap from "post-event treatment" to "regular prevention" and further to "precision prevention." Early identification of disease signs through AI technology can not only avoid agricultural losses caused by large-scale disease spread but also achieve precision pesticide application, reduce the environmental and ecological impacts of chemical pesticides, ensure the quality and safety of agricultural products, and build a truly intelligent crop health management system.

Participants will receive a crop leaf disease image dataset—a large-scale, high-quality synthetic dataset specifically designed for smart agriculture application scenarios. The dataset includes:

- **Large scale:** Covers 30,000 high-definition image samples of 61 common crop diseases, ensuring data diversity.
- **Rich crop types:** Includes major agricultural categories such as food crops,

cash crops, vegetables, and fruits.

- **Comprehensive disease types:** Encompasses major disease categories such as fungal diseases, bacterial diseases, viral diseases, and physiological diseases.

- **Excellent image quality:** All images are collected under standard lighting conditions, including samples of different disease stages and severity levels.

- **Professional annotation:** Precisely annotated by an agricultural expert team, with each image containing multi-dimensional labels such as disease type, severity, and crop variety.

- **Practical scenario data:** Includes images taken in natural field environments, truly reflecting real application scenarios.

(II) Dataset Interpretation

The original dataset consists of two folders:

The training set (AgriculturalDisease_trainingset) and the validation set (AgriculturalDisease_validationset), which contain image files (images) used for training and validation, as well as JSON files labeling image types. The total number of training images is 32,768, and the total number of validation images is 4,992.

The dataset includes 61 categories (classified by "species-disease-severity"), 10 species, 27 diseases (24 of which are divided into general and serious severity levels), and 10 healthy categories. The specific classifications are as follows:

Label id	Label name	Label id	Label name
0	Apple Healthy	31	Pepper Scab (General)
1	Apple Scab (General)	32	Pepper Scab (Serious)
2	Apple Scab (Serious)	33	Potato Healthy
3	Apple Frogeye Spot	34	Potato Early Blight (Fungus, General)
4	Cedar Apple Rust (General)	35	Potato Early Blight (Fungus, Serious)
5	Cedar Apple Rust (Serious)	36	Potato Late Blight (Fungus, General)
6	Cherry Healthy	37	Potato Late Blight (Fungus, Serious)
7	Cherry Powdery Mildew (General)	38	Strawberry Healthy

8	Cherry Powdery Mildew (Serious)	39	Strawberry Scorch (General)
9	Corn Healthy	40	Strawberry Scorch (Serious)
10	Cercospora Zeaemaydis Tehon and Daniels (General)	41	Tomato Healthy
11	Cercospora Zeaemaydis Tehon and Daniels (Serious)	42	Tomato Powdery Mildew (General)
12	Corn Puccinia Polysora (General)	43	Tomato Powdery Mildew (Serious)
13	Corn Puccinia Polysora (Serious)	44	Tomato Bacterial Spot (Bacteria, General)
14	Corn Curvularia Leaf Spot (Fungus, General)	45	Tomato Bacterial Spot (Bacteria, Serious)
15	Corn Curvularia Leaf Spot (Fungus, Serious)	46	Tomato Early Blight (Fungus, General)
16	Maize Dwarf Mosaic Virus	47	Tomato Early Blight (Fungus, Serious)
17	Grape Healthy	48	Tomato Late Blight (Water Mold, General)
18	Grape Black Rot (Fungus, General)	49	Tomato Late Blight (Water Mold, Serious)
19	Grape Black Rot (Fungus, Serious)	50	Tomato Leaf Mold (Fungus, General)
20	Grape Black Measles (Fungus, General)	51	Tomato Leaf Mold (Fungus, Serious)
21	Grape Black Measles (Fungus, Serious)	52	Tomato Target Spot (Bacteria, General)
22	Grape Leaf Blight (Fungus, General)	53	Tomato Target Spot (Bacteria, Serious)
23	Grape Leaf Blight (Fungus, Serious)	54	Tomato Septoria Leaf Spot (Fungus, General)
24	Citrus Healthy	55	Tomato Septoria Leaf Spot (Fungus, Serious)
25	Citrus Greening (General)	56	Tomato Spider Mite Damage (General)
26	Citrus Greening (Serious)	57	Tomato Spider Mite Damage (Serious)
27	Peach Healthy	58	Tomato Yellow Leaf Curl Virus (General)
28	Peach Bacterial Spot (General)	59	Tomato Yellow Leaf Curl Virus (Serious)
29	Peach Bacterial Spot (Serious)	60	Tomato Mosaic Virus
30	Pepper Healthy	-	-

(III) Tasks to be Solved

Task 1: Optimization of Agricultural Disease Image Classification Model

Task Description: Given the dataset of 61 categories of agricultural disease images, participants need to design and train a deep learning model to achieve high-precision disease classification.

Specific Requirements:

1. Use the provided training set (including 61 category folders) for model training.

2. The dataset contains dirty data—images labeled with "duplicate" indicate repeated annotations; data cleaning must be performed first.

3. Model architecture can be freely selected (CNN, Transformer, etc.).

4. Must include data preprocessing and data augmentation strategies.

5. Model parameter count shall not exceed 50M.

6. Training time is limited to 24 hours.

Task 2: Few-Shot Agricultural Disease Recognition

Task Description: Achieve effective classification of 61 categories of agricultural diseases under the extreme condition of only 10 training images per category.

Specific Requirements:

1. Randomly select 10 images per category from the complete dataset as the training set.

2. Technologies such as transfer learning, meta-learning, and data generation can be used.

3. Additional training data is prohibited.

4. Model parameter count shall not exceed 20M.

Task 3: Prediction of Disease Severity Grading

Task Description: Construct a deep learning model using crop leaf images to predict disease severity levels (healthy, mild, moderate, severe).

Specific Requirements:

1. Automatically determine disease severity based on image content (4-class classification task).

2. Independently judge how to implement 4-class classification from labels in the JSON file and group image data accordingly.

3. Output model accuracy, macro-averaged F1-score, and recall rate for each category.

4. Visualize model-focused key regions (e.g., using technologies like Grad-CAM).

Task 4: Multi-Task Joint Learning and Interpretable Diagnosis

Task Description: Construct a multi-task learning system to simultaneously complete disease classification, severity grading, and provide interpretable diagnostic reports.

Specific Requirements:

1. Simultaneously output disease type and severity level.

2. Generate readable diagnostic reports (including key information such as confidence level).

3. Evaluate the synergistic effect of multi-task learning.