Data Collection

Data Set

Date	2-11-2022
Team ID	PNT2022TMID17339
Project Name	Fertilizers Recommendation System for Disease Prediction
Maximum Marks	8mark

Data Collection:

In Data Collection are the important task to build a Project in Artificial Intelligences (AI). In Modern Artificial Intelligences are using the Deep Neural Network to obtain a maximum accuracy compare to the Machine Learning algorithm in Computer vision Field.

In Disease Image Classification we can collect a maximum image per classes to improving the Accuracy of a Deep Learning Model.

Data Source for Crop Leaf Disease Image Classification:

1. PlantDoc:

In plant doc dataset contains 2,598 data points in total across 13 plant species and up to 17 classes of diseases, involving approximately 300 human hours of effort in annotating internet scraped images.

- ➢ PlantDoc Official Paper:
  https://arxiv.org/abs/1911.10317#:~:text=Against%20this%20background%2C%20we%20present,in%20annotating%20internet%20scraped%20images.
- ➢ PlantDoc Download (Kaggle):
  https://www.kaggle.com/datasets/abdulhasibuddin/plant-doc-dataset

2. PlantVillage:

The PlantVillage dataset consists of 54303 healthy and unhealthy leaf images divided into 38 categories by species and disease. Note: The original dataset is not available from the original source (plantvillage.org), therefore we get the unagmented dataset from a paper that used that dataset and republished it

- ➢ PlantVillage Official Paper:
  https://paperswithcode.com/paper/improving-plant-disease-classification-by
- ➢ PlantVillage Download (Kaggle):
  https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset

Apple scab   Cherry powdery mildew   Corn northern leaf blight   Grape black rot   Grape leaf blight   Orange Haunglongbing (Citrus greening)

Peach bacterial spot   Potato early blight   Squash powdery mildew   Strawberry leaf scorch   Tomato early blight   Tomato late blight

Project Dataset:

In those above two dataset are integrate into a single dataset for using our project to obtain the maximum accuracy form the given Dataset.