Data Collection and Preprocessing Phase

Date	9 July 2024
Name	Pratik Prasad Mahadik
Project Title	Greenclassify: Deep Learning-Based Approach For Vegetable Image Classification
Maximum Marks	2 Marks

Data Quality Report Template

The Vegetable Image Dataset on Kaggle contains 21,000 high-resolution (224x224) JPG images of 15 different vegetable types, with 1,400 images per class. The dataset is balanced and publicly available, suitable for training and validating deep learning models. It is recommended to split the data into 70% for training, 15% for testing, and 15% for validation. Quality considerations include ensuring varied image conditions and accurate labeling for effective model training.

Data Source	Data Quality Issue	Severity	Resolution Plan
Vegetable Image Dataset	Inconsistent dimensions of images in the dataset. Need to standardize the image dimensions to either 224x224 or 299x299.	High	Resize all images to 224x224 using the flow_from_directory method in the Keras ImageDataGenerator
	Images need normalization.	Moderate	Normalize the pixel values to a standard range (e.g., [0, 1]) using ImageDataGenerator's rescale parameter