NitroGrowth - Smart Plant Nitrogen Estimation System

A PROJECT REPORT

Submitted by

Rishab sharma -20BCS1849

Samvart Verma-20BCS5345

Rahul Kadyan-20BCS1086

Chirgah Rohilla-20BCS5404

Sharad singh-20BCS1861

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE ENGINEERING

Chandigarh University

May 2023

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

Phase-2

Data Collection and Preprocessing

Introduction

Phase 2 of our project is dedicated to the critical tasks of data collection and preprocessing. These activities are foundational to acquiring the necessary data to train and validate our nitrogen content estimation algorithms. The objectives in this phase are threefold: firstly, to gather a diverse range of plant images along with their associated nitrogen content measurements; secondly, to enhance the quality of these images through preprocessing techniques; and thirdly, to ensure data consistency and reliability through data standardization and quality assurance.

Objectives

Data Acquisition:

The first objective is to collect an extensive dataset of plant images and their corresponding nitrogen content measurements. This dataset must be diverse and representative to ensure the robustness and adaptability of our estimation system.

Image Preprocessing:

The second objective involves enhancing the quality of the collected images. This is achieved through various techniques such as contrast adjustment, noise reduction, and image sharpening. High-quality images are vital for accurate nitrogen content estimation.

Data Standardization and Quality Assurance: The third objective focuses on data standardization and quality assurance. Standardizing data formats ensures compatibility with our algorithm development pipeline. Simultaneously, we implement rigorous quality control measures to validate data accuracy and reliability, identifying and addressing outliers and data inconsistencies.

Data Collection

Sources of Data:

To meet our data acquisition objectives, we employ two primary sources:

Drones: Drones equipped with high-resolution cameras are used to capture aerial images of agricultural fields. These images offer a comprehensive view of various crop types and growth stages across larger areas, allowing for efficient assessment of plant health and nitrogen content.

Smartphones: Data collection also involves smartphones with built-in cameras. This approach enables us to gather localized and detailed data, focusing on specific plant growth stages, disease symptoms, and other critical factors. Smartphone data complements drone imagery by providing finer detail and context.

Plant Types and Growth Stages

We ensure diversity in our dataset by collecting images of various plant types (e.g., wheat, corn, soybeans) at different growth stages (e.g., seedling, vegetative, reproductive). This diversity is essential for training our algorithms to handle the complexities and variations encountered in real-world agricultural scenarios.

Nitrogen Content Data:

Concurrently with image collection, we gather nitrogen content data through laboratory measurements and field analysis. These measurements serve as ground truth data, allowing us to validate the accuracy of our nitrogen content estimation algorithms.

Data Preprocessing

Image Preprocessing:

Image Enhancement:
To improve image quality, we employ a suite of image enhancement techniques. These techniques include adjusting contrast, reducing noise, and sharpening images. Enhanced images provide a clearer and more informative basis for nitrogen content estimation.

Image Registration:
Proper spatial alignment of images is crucial to facilitate accurate feature extraction and analysis. Image registration ensures that images are correctly aligned, preventing misalignment-induced errors.

Data Augmentation:
Data augmentation techniques, such as image rotation, scaling, and flipping, are used to expand our training dataset. Augmentation enhances the robustness of our algorithms by exposing them to various perspectives and conditions.

Data Standardization

Data Formatting:
Standardizing data formats is essential to ensure compatibility with our algorithm development pipeline. Consistency in data formatting simplifies data integration and analysis, streamlining the algorithm development process.

Missing Data Handling:
Robust strategies are developed to address missing or incomplete data points. Proper handling of missing data ensures that our dataset remains complete and reliable, reducing potential biases in our estimations.

Quality Control

Outlier Detection: Identifying and handling outliers is paramount for maintaining the integrity of our dataset. Outliers can significantly impact algorithm training and analysis. We employ statistical methods and visual inspection to identify and address outliers in both image and nitrogen content data.

Data Validation:
Regular data validation checks are conducted to ensure the accuracy and consistency of the collected data. Data validation processes help identify and rectify data inconsistencies, ensuring that our dataset is of the highest quality.

Timeline

To provide a structured framework for Phase 2 activities, we have established the following timeline:

Data Collection: [Start Date] - [End Date]
Image Preprocessing: [Start Date] - [End Date]
Data Formatting: [Start Date] - [End Date]
Quality Control: [Start Date] - [End Date]

Budget Allocation:

We have allocated a portion of the project budget to Phase 2 activities, encompassing expenses related to data collection equipment, data storage infrastructure, and any necessary software tools for data preprocessing. Proper budget allocation ensures that the data collection and preprocessing phases are adequately funded.

Conclusion

Phase 2, with its focus on data collection and preprocessing, represents a pivotal juncture in the development of the Plant Nitrogen Content Estimation System. The quality, diversity, and reliability of the data gathered and processed during this phase serve as the bedrock upon which we build and train our nitrogen content estimation algorithms. The success of our system ultimately hinges on the meticulous execution of these critical tasks.