LEAF DISEASE CLASSIFICATION

Problem Statement and Application

Crop yield and health are vital factors in the agricultural industry. It is essential to address the challenges especially leaf disease as it impacts the quality and quantity of the produced yield, resulting in financial losses and food shortages. Plant diseases have traditionally been identified and diagnosed by human examination, which is labor-intensive, time-consuming, and prone to mistakes. Recent developments in computer vision and deep learning provide intriguing means of automating disease detection procedures, hence transforming agricultural methods. Our objective is to develop a reliable, efficient automated plant disease detection system that enhances the sustainability and resilience of global food systems.

Plant diseases are difficult to classify because of their wide range of symptoms, which are regulated by temperature and soil conditions, among other environmental variables. It is difficult to obtain high-quality training data, as it requires several datasets including images of diseased plants.

Our major objective is to efficiently identify plant diseases from leaf images using convolutional neural networks (CNNs). Gaining knowledge about how environmental factors affect illness manifestation will strengthen the robustness of the model. Additionally, we aim to analyze the impact of dataset size and other image factors like dimension, color etc on model effectiveness. We evaluate deep learning architectures and methodologies with the goal of optimizing performance. Our goal is to research, evaluate, and compare several models in order to determine which is best for classifying leaf diseases. This information will then be useful for a variety of applications, including medical ones.

Image Dataset Selection

Info	**PlantVillage**	**Mendeley**	**Rice Leaf**
Classes	38	22	4
Images	30.4K	24.8K	5.9K
Dimension	256*256	>400*400	300 * 300
Format	.jpg	.jpg	.jpg

PlantVillage: This dataset consists of various plant leaf disease images in color, grayscale, and segmented forms[1]. **Mendeley**: This dataset consists of raw data and segmented images of different plants, categorized into 22 classes[2]. **Rice Leaf**: This dataset consists four kinds of Rice leaf diseases i.e. Bacterial blight, Blast, Brown Spot and Tungro.[3].

Possible Methodology

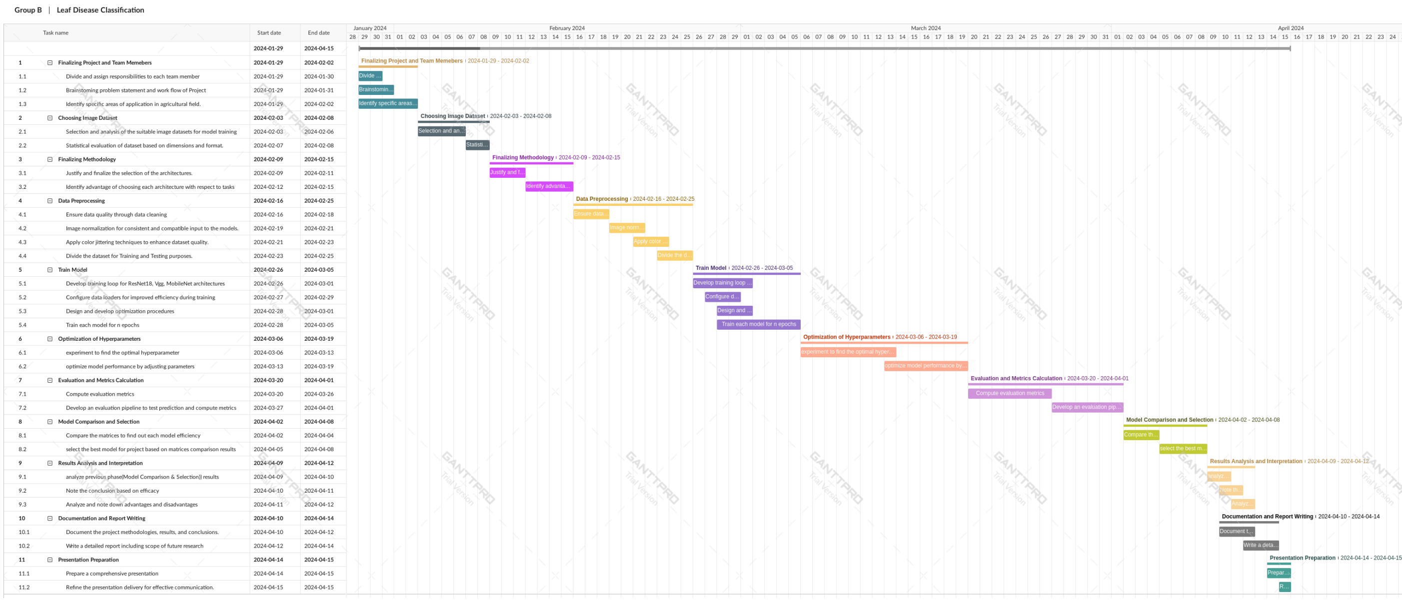
Pre-Processing: Resize images to the model's input size and normalize pixel values for consistency. Apply data augmentation techniques like random cropping, flipping, and rotation to enhance dataset diversity. Apply color space conversion like RGB, HSV, grayscale which optimizes image representation. Categorical labels may be encoded as one-hot vectors. Split the dataset into training, validation, and testing sets for robust model development and evaluation.

Model Training: For the model training, ResNet18 can efficiently collect complex characteristics inside pictures while minimizing the vanishing gradient problem. It is well-known for its deep residual learning capacity. The computational efficiency of MobileNetV2, which is attained using depthwise separable convolutions, is well known. This design minimizes the amount of calculations and parameters needed, which makes it suitable for deployment on embedded and mobile devices without compromising accuracy. VGGNet19 uses stack of multiple convolutional layers with small filter sizes, which efficiently captures hierarchical features in images. Each models will be trained on 3 different datasets, resulting in a total of 9 models. Furthermore, for transfer learning, two of the trained models will facilitate the transfer of information and acquired representations by acting as base models on datasets different from their own. Through periodic evaluations, key hyperparameters including learning rates, batch sizes, and epochs will be tuned.

Model Assessment: We'll evaluate models with multi-class metrics like accuracy, precision, recall, and F1 score. Confusion matrices will show class-wise performance. TSNE visualization will aid in understanding class separability. We aim for high scores across metrics, ensuring accurate disease classification and guiding agricultural practices for improved crop management.

Applications of Derived results: Our research provides information on the effectiveness of particular models, the effects of preprocessing methods, hyperparameter tuning, and transfer learning in the classification of leaf diseases. With this thorough knowledge, scientists may enhance their methods by experimenting with new preprocessing techniques, adjusting hyperparameters, and using transfer learning to better adapt their models to a variety of datasets. By combining these methods, scientists may create machine learning frameworks that are more resilient and flexible for a range of plant pathology applications, spurring creativity and progress in the field of agricultural research.

Gantt Chart


Group B | Leaf Disease Classification
Task name
Start date
End date
January 2024
February 2024
March 2024
April 2024
2024-01-29 2024-04-15
1 Finalizing Project and Team Memebers 2024-01-29 2024-02-02
1.1 Divide and assign responsibilities to each team member 2024-01-29 2024-01-30
1.2 Brainstoming problem statement and work flow of Project 2024-01-29 2024-01-31
1.3 Identify specific areas of application in agricultural field. 2024-01-29 2024-02-02
2 Choosing Image Dataset 2024-02-03 2024-02-08
2.1 Selection and analysis of the suitable image datasets for model training 2024-02-03 2024-02-06
2.2 Statistical evaluation of dataset based on dimensions and format. 2024-02-07 2024-02-08
3 Finalizing Methodology 2024-02-09 2024-02-15
3.1 Justify and finalize the selection of the architectures. 2024-02-09 2024-02-11
3.2 Identify advantage of choosing each architecture with respect to tasks 2024-02-12 2024-02-15
4 Data Preprocessing 2024-02-16 2024-02-25
4.1 Ensure data quality through data cleaning 2024-02-16 2024-02-18
4.2 Image normalization for consistent and compatible input to the models. 2024-02-19 2024-02-21
4.3 Apply color jittering techniques to enhance dataset quality. 2024-02-21 2024-02-23
4.4 Divide the dataset for Training and Testing purposes. 2024-02-23 2024-02-25
5 Train Model 2024-02-26 2024-03-05
5.1 Develop training loop for ResNet18, Vgg, MobileNet architectures 2024-02-26 2024-03-01
5.2 Configure data loaders for improved efficiency during training 2024-02-27 2024-02-29
5.3 Design and develop optimization procedures 2024-02-28 2024-03-01
5.4 Train each model for n epochs 2024-02-28 2024-03-05
6 Optimization of Hyperparameters 2024-03-06 2024-03-19
6.1 experiment to find the optimal hyperparameter 2024-03-06 2024-03-13
6.2 optimize model performance by adjusting parameters 2024-03-13 2024-03-19
7 Evaluation and Metrics Calculation 2024-03-20 2024-04-01
7.1 Compute evaluation metrics 2024-03-20 2024-03-26
7.2 Develop an evaluation pipeline to test prediction and compute metrics 2024-03-27 2024-04-01
8 Model Comparison and Selection 2024-04-02 2024-04-08
8.1 Compare the matrices to find out each model efficiency 2024-04-02 2024-04-04
8.2 select the best model for project based on matrices comparison results 2024-04-05 2024-04-08
9 Results Analysis and Interpretation 2024-04-09 2024-04-12
9.1 analyze previous phase(Model Comparison & Selection)) results 2024-04-09 2024-04-10
9.2 Note the conclusion based on efficacy 2024-04-10 2024-04-11
9.3 Analyze and note down advantages and disadvantages 2024-04-11 2024-04-12
10 Documentation and Report Writing 2024-04-10 2024-04-14
10.1 Document the project methodologies, results, and conclusions. 2024-04-10 2024-04-12
10.2 Write a detailed report including scope of future research 2024-04-12 2024-04-14
11 Presentation Preparation 2024-04-14 2024-04-15
11.1 Prepare a comprehensive presentation 2024-04-14 2024-04-15
11.2 Refine the presentation delivery for effective communication. 2024-04-15 2024-04-15

Bibliography

1. SP Mohanty. PlantVillage. https://github.com/spMohanty/PlantVillage-Dataset

2. Patrick Mensah Kwabena, Vivian Akoto-Adjepong, Kwabena Adu, Mighty Abra Ayidzoe, Elvis Asare Bediako, Owusu Nyarko-Boateng, Samuel Boateng, Esther Fobi Donkor, Faiza Umar Bawah, Nicodemus Songose Awarayi, Peter Nimbe, Isaac Kofi Nti, Muntala Abdulai, Remember Roger Adjei, Michael Opoku. Dataset for Crop Pest and Disease Detection. DOI: 0.17632/bwh3zbpkpv.1. https://data.mendeley.com/datasets/bwh3zbpkpv/1

3. Prabira Kumar Sethy. Rice Leaf Disease Image Samples. DOI: 10.17632/fwcj7stb8r.1. https://data.mendeley.com/datasets/fwcj7stb8r/1

4. P. Sirenjeevi, J. Madhu Karthick, K. Agalya, R. Srikanth, T. Elangovan, and R. Nareshkumar. Leaf Disease Identification using ResNet. DOI: 10.1109/ICECONF57129.2023.10083963. https://ieeexplore.ieee.org/document/10083963

5. Elhoucine Elfatimi, Recep Eryigit, Lahcen Elfatimi. Beans Leaf Diseases Classification Using MobileNet Models. DOI: 10.1109/ACCESS.2022.3142817. https://ieeexplore.ieee.org/document/9681085

6. Dr. Thanh-Hai Nguyen, Dr. Nghia Thanh Nguyen, Mr. Ba-Viet Ngo. A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease. https://www.mdpi.com/2624-7402/4/4/56

7. Jinzhu Lu, LijuanTan, Huanyu Jiang. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. https://doi.org/10.3390/agriculture11080707

8. Wubetu Barud Demilie. Plant disease detection and classification techniques: a comparative study of the performances. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00863-9

9. Plant Disease Detection using AI based VGG-16 Model. https://thesai.org/Downloads/Volume13No4/Paper_84-Plant_Disease_Detection_using_AI_based_VGG_16_Model.pdf