frontier tech leaders

A new generation of tech specialists

Plant Disease Detection Using Deep Learning

Outline

- Concept note and implementation plan:
    - Background
    - Objectives
    - SDG Relation
- Data
    - Data Collection
    - Exploratory Data Analysis (EDA) and Feature Engineering
- Model Selection and Training
    - Model Evaluation and Hyperparameter Tuning
    - Model Refinement and Testing
- Results
- Deployment
- Future Work

Concept Note and Implementation

Background

- The "Plant Disease Detection Using Deep Learning" project aims to boost agricultural productivity by using AI to detect plant diseases early. Supporting SDGs 2, 3, and 15, it addresses food security, health, and sustainable agriculture. The automated system identifies diseases from leaf images, minimizing crop losses and reducing pesticide use, benefiting regions like Ethiopia.

- Developing countries' agriculture faces challenges like climate change, pests, and plant diseases. Traditional detection methods are slow and error prone. Deep learning, especially CNNs, provides a high-accuracy, automated alternative, suitable for large-scale deployment.
- The project addresses significant agricultural challenges, including slow and error-prone traditional disease detection methods, leading to extensive crop losses. By using deep learning for accurate, automated detection, it enhances crop yields and reduces pesticide dependency. This supports SDGs 2, 3, and 15, promoting food security, health, and sustainability

Objectives

- Develop a deep learning model capable of accurately detecting various plant diseases from leaf images.
- Create a scalable and robust system that can be accessible on portable devices for real-time disease detection in the field.
- Enhance agricultural productivity by enabling early intervention and reducing crop losses.
- Promote environmentally sustainable farming practices by minimizing the use of chemical pesticides.

SDG Relation

SDG 2: Zero Hunger

- Enhance agricultural productivity
- Minimize crop losses due to diseases

2. SDG 3: Good Health and Well-being

- Reduce dependency on chemical pesticides
- Protect health of farmers and consumers

3. SDG 15: Life on Land

- Promote sustainable agricultural practices
- Reduce overuse of pesticides

Data

Data Collection

- Source: The primary dataset for this project is the PlantVillage dataset, which contains 55,448 images of healthy and diseased plant leaves across 39 crop categories. This dataset is known for its diversity and quality, making it ideal for training robust deep-learning models.
- Data Augmentation: To simulate various environmental conditions and increase dataset variability, techniques such as scaling, shifting, rotation and flipping were applied.
- Image Standardization: All images were resized to ensure uniformity.
- Image Scaling: To make sure it's easy for the learning process to go smoothly, the values of the RGB pixel values were scaled.
- Checked for any missing values or corrupted images in the dataset. Any incomplete data entries were removed or replaced.
- Outliers: Inspected the dataset for outliers, such as images with incorrect labels or abnormal features, and cleaned these entries to prevent skewed results

Exploratory Data Analysis (EDA) and Feature Engineering

- Image Augmentation: Applied transformations like rotation, scaling, and flipping to create more training examples and improve model robustness.

- Rationale: These features were engineered to provide the model with diverse and informative inputs that enhance its ability to distinguish between healthy and diseased leaves under various conditions.

- Scaling: Resized images to a uniform size of 240x240 pixels.
- Normalization: Normalized pixel values to the range [0, 1] to facilitate faster convergence during training.

Model

Model Selection and Training

- Rationale: CNNs are particularly well-suited for image classification tasks due to their ability to automatically learn spatial hierarchies of features from images. They have shown high accuracy in similar tasks and can efficiently handle the complexity of plant disease detection.

Strengths:

- Excellent at feature extraction and pattern recognition in images.
- Robust to variations in image quality and background.
- Scalable and adaptable to different datasets.

Weaknesses:

- Requires a large amount of labeled data for training.
- Computationally intensive, requiring significant processing power.
- Details on training the model
- Hyperparameters: Learning rate, batch size, number of epochs, dropout rate.
- Cross-Validation: Used k-fold cross-validation to ensure the model's robustness and generalization capability.

Model Evaluation and Hyperparameter Tuning

- Accuracy and loss were the main metric used. The initial model achieved an accuracy of around 78% on the test (was named validation in our case) set.
- Confusion matrix showed misclassifications in some classes.
- Conducted multiple training runs with smaller epochs to find the optimal combination of hyperparameters.
- Key insights included the importance of a smaller learning rate and using Adam optimizer instead of RMS prop which resulted in a lower loss and higher accuracy.
- Impact: Improved accuracy by 12%. Enhanced precision and recall metrics, particularly for classes with less data.

Model Refinement and Testing

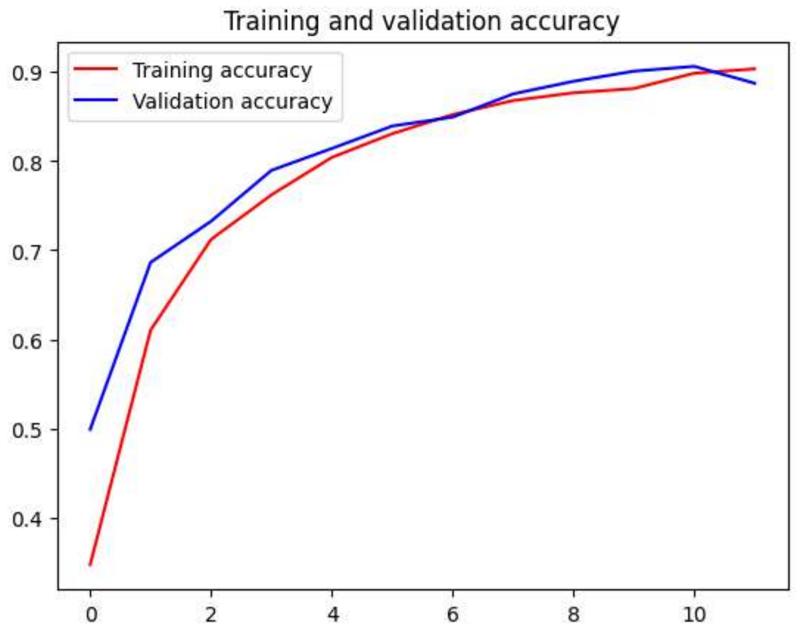
- Our model refinement phase focused on iterative improvements using hyperparameter tuning, algorithm adjustments, and feature selection to enhance performance and generalization on unseen data. The goal was to achieve optimal model performance.
- Hyperparameter Tuning: Adjusted learning rates, batch sizes, and epoch number.
- Algorithm Variations: Explored different CNN architectures starting with smaller architectures with up to 16 convolutional filters and then scaled up to compared performance.
- Overview: The test submission phase involves preparing the refined model for final evaluation on a separate test dataset. This phase ensures the model's robustness and readiness for real-world deployment.
- Metrics: Manual Accuracy, precision, recall, F1-score on the test dataset after training.
- Accuracy on test data: 90% using validation generator (used as test set)
- Results indicated strong generalization capability in the latest training run before diminishing returns after the 12th epoch.

Result

Evaluation Results

Results:

- Accuracy on test data: 90% using validation generator (used as test set)
- Results indicated strong generalization capability in the latest training run before diminishing returns after the 12th epoch



Deployment

Deployment

- The deployment phase involves making the trained CNN for plant disease detection accessible, including model serialization, serving, API integration, security, and monitoring. The goal is a simple, functional deployment for presentation purposes.
- The trained model is serialized using Keras and stored in the HDF5 format (.h5) for efficient storage and easy loading. The model size is 2 MB, facilitating a smooth loading process.

Conclusion and Future work

Conclusion

- The model refinement and test submission phases were crucial in enhancing the performance and reliability of the plant disease detection system. Through iterative improvements, hyperparameter tuning, and robust testing, the model achieved an accuracy of 90% on the test dataset, demonstrating strong generalization capabilities. Challenges such as data imbalance and model complexity were addressed, resulting in a streamlined and effective model ready for real-world deployment.

Thank you!

frontier
tech
leaders