© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 05 | October - December 2023
Refereed & Peer Reviewed
DOI : https://doi.org/10.36676/jrps.2023-v14i5-09

AI-Powered Plant Disease Classification: Innovating for Sustainable Agriculture

Arnav Bansal
arnavb2024@gmail.com

Abstract

This research explores the development and application of artificial intelligence (AI) models in Python for plant disease classification. Using a large training dataset with over 50,000 images representing various plant conditions, the study highlights the effectiveness of AI and computer vision by achieving approximately 86% accuracy in correctly identifying the plant disease. It demonstrates a balanced approach to maintaining accuracy while avoiding overfitting, underscoring AI's potential in agriculture.

Introduction

Plant diseases pose a significant threat to global agriculture, jeopardizing food security and economic stability. As defined by the Food and Agriculture Organization (FAO), plant diseases encompass a wide range of harmful conditions caused by pathogens, environmental factors, and genetic abnormalities that impair the growth and health of crops. These diseases can lead to devastating consequences, including reduced crop yields, increased production costs, and a higher risk of food shortages. The critical nature of this issue is underscored by the fact that plant diseases are estimated to cause an annual loss of up to 40% of global crop production.

The importance of addressing the problem of plant diseases cannot be overstated. In a world where the demand for food continues to rise due to population growth, achieving sustainable agricultural practices is paramount. The loss of crop yield due to plant diseases exacerbates the challenges of feeding a growing global population. Furthermore, the economic impact of these diseases is staggering, with billions of dollars lost each year in agricultural revenue and increased expenditure on pesticides and other disease management measures.

Early detection and effective management of plant diseases are crucial to mitigating their impact. Identifying diseases at their onset allows for timely intervention, reducing the spread of pathogens and minimizing yield losses. Therefore, the development of accurate and efficient disease detection methods is imperative for sustainable agriculture. Currently, plant diseases are often detected by trained professionals. This can be time consuming and resource intensive.

This research paper aims to address this pressing issue by leveraging the power of machine learning and artificial intelligence to classify plant diseases. We propose to explore and compare various AI models to determine the most effective approach for early disease detection. By harnessing machine learning and deep learning techniques, we aim to provide a reliable tool for farmers and agricultural stakeholders to accurately identify plant diseases. With the ResNet50 model, we were able to achieve 86% accuracy. The potential benefits of this research include a significant reduction in crop yield losses, lower production costs, and a more sustainable and resilient agricultural system. In doing so, we contribute to the broader effort of ensuring food

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 05 | October - December 2023
Refereed & Peer Reviewed
DOI : https://doi.org/10.36676/jrps.2023-v14i5-09

security and sustainable agriculture in the face of mounting global challenges.

Dataset and Preprocessing

The dataset was taken from PlantVillage, an online platform. There are over 50,000 images of healthy and infectious diseases affecting different plants. This dataset contains 38 different types of healthy/infected plants. Here are the types:

Plant Type	Disease Conditions
Apple	Apple_scab, Black_rot, Cedar_apple_rust, Healthy
Blueberry	Healthy
Cherry (including sour)	Powdery_mildew, Healthy
Corn (maize)	Cercospora_leaf_spot Gray_leaf_spot, Common_rust, Northern_Leaf_Blight, Healthy
Grape	Black_rot, Esca (Black_Measles), Leaf_blight (Isariopsis_Leaf_Spot), Healthy
Orange	Haunglongbing (Citrus_greening)
Peach	Bacterial_spot, Healthy
Pepper, bell	Bacterial_spot, Healthy
Potato	Early_blight, Late_blight, Healthy
Raspberry	Healthy
Soybean	Healthy
Squash	Powdery_mildew
Strawberry	Leaf_scorch, Healthy

Table 1: Plant Diseases and Health Conditions

The first step was splitting our data into train, test, and valid sets (80% train, 10% test, 10% valid). The next step was flattening the images, which is our x variable. This means that each row in the datasets is converted into a single vector. This is necessary for linear regression modeling. The last step in the preprocessing of the data was converting the one-hot encoded target variables y_train, y_valid, and y_test into their corresponding class labels, which were the types of healthy/infected plants.

Model Development and Approach

Before developing our model, we ran tests with baseline models to contextualize our final results and measure our accuracy. For all of our baseline models, we used scikit-learn implementations.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 05 | October - December 2023
Refereed & Peer Reviewed
DOI : https://doi.org/10.36676/jrps.2023-v14i5-09

We tested both a standard linear regression and multi layer perceptron (MLP) classifier. Between the two, linear regression performed best with an accuracy of roughly 69.5%. This might be explained by the relatively small amount of data and few iterations of training.

Baseline Model	**Accuracy**
Sklearn LinearRegression	0.695
Sklearn MLPClassifier	0.523

Table 2: Baseline Model Results

After establishing baseline performance with simpler models, After establishing baseline performance with simpler models, we explored ResNet, or Residual Network, which is a type of advanced convolutional neural network used in deep learning. It's known for its "skip connections" which allow it to skip certain layers, helping to solve the vanishing gradient problem in deep networks. This makes it possible to build much deeper networks, improving performance on complex tasks like image recognition. ResNet variants, like ResNet-50, are named based on the number of layers they contain. Two of pre-trained models based on ResNet architecture were evaluated to determine the most effective for the complex task of plant disease classification.

In our research using ResNet models for plant disease classification, the choice of optimizer, Adam or SGD, significantly impacts model performance. Adam, known for efficiently handling large datasets and complex models like ResNet50, aids in faster convergence with automatic learning rate adjustments. Meanwhile, SGD, a more traditional but simpler optimizer, may require careful tuning but can be effective, especially with techniques like momentum. In this case, Adam may have worked better because it is well-suited for dealing with large and complex datasets like those involved in plant disease classification. Its ability to automatically adjust the learning rate helps in navigating the intricacies of high-dimensional data more effectively than SGD. This leads to faster convergence and more robust performance in complex models like ResNet50, which is essential for accurately classifying the diverse and nuanced patterns present in plant disease images. Multiple iterations were run and data was collected in Table 3.

Model	Parameters	Train Acc	Valid Acc	Train Loss	Valid Loss
ResNet 50	ADAM LR = 0.0001, CrossEntropy, 30 epochs	**0.8536**	**0.8322**	**0.4591**	**0.5326**
ResNet 18	ADAM LR = 0.0001 CrossEntropy, 30 epochs	**0.7822**	**0.7711**	**0.6741**	**0.7200**
ResNet50	SGD, CrossEntropy, 30 epochs	**0.8363**	**0.8203**	**4.6081**	**7.1742**
ResNet 50	ADAM LR = 0.0001 CrossEntropy, 50 epochs	**0.8660**	**0.8448**	**0.4096**	**0.4857**

Table 3: Model Performance Comparison

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 05 | October - December 2023
Refereed & Peer Reviewed
DOI : https://doi.org/10.36676/jrps.2023-v14i5-09

ResNet50 emerged as the preferred choice due to its advanced architecture and proven ability to handle complex datasets. ResNet50, a deep convolutional neural network, is known for its depth and ability to learn from a large number of parameters, making it particularly suited for intricate image recognition tasks like identifying various plant diseases. Its selection was based on a comparative analysis which showed that ResNet50 outperformed other models, including ResNet18, in terms of accuracy and reliability in dealing with the dataset's complexity, as shown in Table 3.

Discussion and Learning Outcomes

Accuracy and Loss were analyzed by creating graphs to monitor the model's performance over different epochs, as illustrated in Figures 1 and 2. After evaluating various models, the ResNet50 model equipped with the ADAM optimizer and trained for 50 epochs achieved the highest accuracy and lowest loss. The final accuracy on the test set was 85.73% with a loss of 0.4382. Graphs depicting the accuracy and loss for the training and validation sets indicated that the model did not overfit the data. However, a potential for overfitting was noted if the training continued for more epochs, as suggested by the widening gap in accuracy between the training and validation results, evident in Figure 2.

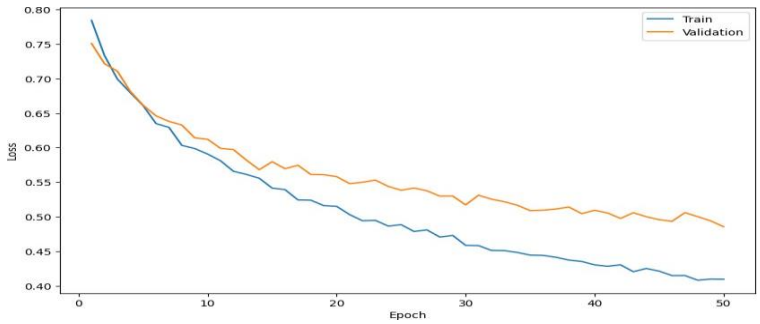

Fig 1. Loss over Epochs

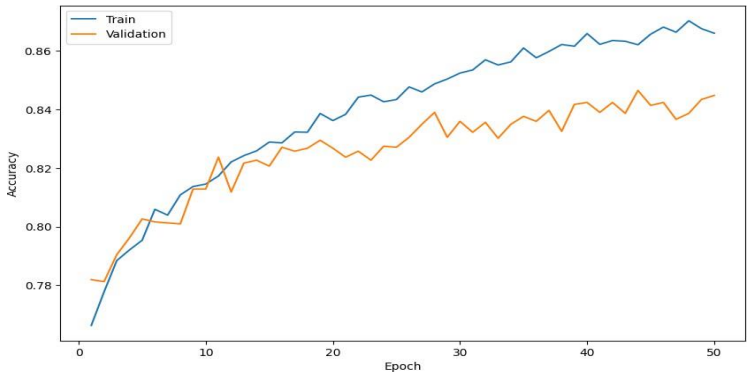

Fig. 2 Accuracy over Epochs

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 05 | October - December 2023
Refereed & Peer Reviewed
DOI : https://doi.org/10.36676/jrps.2023-v14i5-09

Conclusion and Future Work

The research successfully demonstrated the application of advanced AI techniques, particularly using the ResNet50 model, in the classification of plant diseases. The high accuracy achieved underscores the potential of AI in revolutionizing agricultural practices, offering a valuable tool for disease identification and management. The study not only contributes to the field of sustainable agriculture but also opens avenues for further research in applying AI to environmental and agricultural challenges. The findings hold promise for developing practical solutions, such as a user-friendly diagnostic tool, to aid farmers and researchers in early disease detection and management, ultimately leading to healthier crops and more efficient farming practices.

For future work, there are several avenues for improvement and expansion. The development of a user-friendly website where individuals can upload images of plants for instant disease diagnosis will make the model more accessible and practical for real-world applications. Additionally, exploring more advanced neural network architectures and fine-tuning the model could further increase accuracy, potentially exceeding the 90% threshold. Collaborations with agricultural experts and integration of regional disease data can enhance the model's applicability across different geographical areas. Continuous updates to the dataset, incorporating a wider variety of plant diseases, will also ensure the model remains effective and relevant in changing agricultural environments.

References

1. Abdallah Ali Dev. "PlantVillage Dataset." Kaggle, n.d., https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset.
2. He, Kaiming, et al. "Deep Residual Learning for Image Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
3. Kingma, Diederik P., and Jimmy Ba. "Adam: A Method for Stochastic Optimization." arXiv preprint arXiv:1412.6980, 2014.
4. Bottou, Léon. "Large-Scale Machine Learning with Stochastic Gradient Descent." Proceedings of COMPSTAT'2010, Physica-Verlag HD, 2010, pp. 177-186.
5. Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using Deep Learning for Image-Based Plant Disease Detection." Frontiers in Plant Science, vol. 7, 2016, p. 1419, Frontiers.
6. Plant Methods Editors. "Plant Diseases and Pests Detection Based on Deep Learning: A Review." Plant Methods, vol. 17, no. 22, 2021.