International Journal of Pathology and Drugs, 2023,
Vol. 01, Iss. 01, June-Nov, S. No. 001, pp. 1-8
ISSN (Online): 2584-1971

Plant Disease Diagnosis and Classification using Deep Learning

Mohd Furqan[1], Sheenu Rizvi[2], Pawan Singh[3]

*[1](Student, Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Lucknow, Uttar Pradesh, India)

[2](Assistant Professor, Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Lucknow, Uttar Pradesh, India)

[3](Associate Professor, Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Lucknow, Uttar Pradesh, India)

[1]furqan3122@gmail.com, [2]sheenu_r@hotmail.com, [3]pawansingh51279@gmail.com

Abstract

Plant disease diagnosis and classification is an important task in agriculture as it helps in early identification and control of plant diseases, ultimately reducing crop loss and improving food security. In recent years, with the improvements in computer vision and machine learning, researchers have developed various techniques for the automated characterization and identification of plant diseases using images. This paper provides an overview of modern technology used for plant disease diagnosis and classification, including classification, image preprocessing, and feature extraction methods. Additionally, this paper highlights the challenges and future directions in this field, such as improving the accuracy of disease detection. In this paper, the model is trained and tested using a convolutional neural network (CNN). The Plant Village dataset is used to determine the precision of the model. In this paper, 18275 images were taken from the dataset belonging to 17 different classes. The dataset is split into three folders – train, validate and test. Images are distributed as follows: 70% are supplied to the train folder, 10% to the validate folder, and 20% to the test folder.

Keywords

Agriculture, Classification, Diagnosis, Diseases, Learning, Plants, and Training.

*Corresponding Author	How to Cite this Article	To browse
Mohd Furqan, Student, Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Lucknow, Uttar Pradesh, India. Email : furqan3122@gmail.com	Furkan M, Rizvi S, Singh P. Plant Disease Diagnosis and Classification using Deep Learning. Int. J. Pathol. Drugs. 2023;1(1):1-8. **DOI:** https://doi.org/10.54060/pd.2023.1	

Received	Accepted	Online First	Published
2023-06-30	2023-11-23	2023-11-25	2023-12-05

Funding	Ethical Approval
Nil	Nil

1. Introduction

Plant diseases are a substantial threat to agriculture and food safety, causing yield loss and economic damages worldwide [1]. The World Health Organization estimates that around 20% of the world food production is lost to plant diseases each year [2]. The efficient management and prevention of plant diseases, the reduction of crop loss risk, and the increase in food output all depend on the early diagnosis and precise categorization of plant diseases.

The majority of the people of India depend on agriculture, making it an agricultural nation. Agricultural research seeks to boost output and food quality while reducing costs and increasing income. Agrochemicals, seeds, and soil interact in a complex way to produce the agricultural production system [3]. Fruits and vegetables rank as the most significant agricultural goods. To obtain more useful products, a product quality control is basically necessary. Many studies indicate that plant diseases may cause a decline in the level of agricultural goods quality. The normal state of the plant is impaired by diseases, which change or stop important processes including photosynthesis, transpiration, pollination, fertilization, germination, etc.

In our study, we have selected some of the plants from the Plant Village dataset [4], like Apple, Corn, Grape, Potato and Tomato. The images of these plants are then put into the Convolutional Neural Network for testing, training, and validation purposes. The training, validation and testing accuracy of the model came out to be 97.12%, 94.73% and 94.13%.

The rest of the paper is organized as follows. The next section materials and methods describe the process of data collection, pre-processing data preparation, model development, evaluation, and deployment. Section 3 represents Result and Discussion which explains the analysis and summarizes the results. In the last section, the concluding remarks are provided.

2. Materials and Methods

Plant disease diagnosis with Deep Learning is a multi-step process. The initial step is to collect photographs of healthy and deceased plants. The photos must next be preprocessed to improve their quality and remove any noise or extraneous data. The next stage is to use techniques such as Convolutional Neural Networks to extract significant characteristics from the preprocessed photos (CNNs).

For object identification and classification with image databases, CNN models are the best choice [5]. Despite the benefits of CNNs, there are still limitations, such as the lengthy training time and the need for big datasets. Deep CNN models are required to extract low-level and complicated characteristics from images, increasing the complexity of model training [6]. The steps involved in Plant disease diagnosis and classifications using deep learning are:

2.1. Data Collection

Data collection is a crucial step in the diagnosis of plant disease using deep learning. Collecting a diverse and representative dataset is essential for training an accurate and effective model. To ensure a high-quality dataset, images of both healthy plants and plants infected with different types of diseases is collected from Plant Village Dataset from Kaggle [4]. Plant Village dataset contains pictures of various plant types and different stages of disease, and images are captured from various parts of the plant. It is important to label the images with the corresponding plant species and disease type to enable supervised learning. By carefully planning and executing the data collection process, the resulting dataset will serve as a strong basis for training an accurate and effective deep learning model for plant disease diagnosis. All the images used in this dataset are of 256 x 256 pixels.

2.2. Data Pre-processing

Pre-processing data is a crucial step in utilizing deep learning to detect plant diseases. The purpose of pre-processing is to clean and modify raw data so that it can be analyzed. Following are some common data pre-processing approaches for plant disease detection:

2.2.1. Image Resizing and Rescale

Image resizing and rescaling are crucial techniques in data pre-processing for plant disease detection using deep learning [7]. Resizing images to a standard size and rescaling them to focus on the plant can improve the accuracy and efficiency of feature extraction, reduce computation time, and boost the size of the dataset. By resizing the images to a standard size, it ensures that all images have the same dimensions, which is necessary for some deep learning algorithms to work effectively. Image Resizing and Rescaling is done by using Sequential function of keras API of Tensor flow.

```
resize_and_rescale = tf.keras.Sequential([
    layers.experimental.preprocessing.Resizing(IMAGE_SIZE, IMAGE_SIZE),
    layers.experimental.preprocessing.Rescaling(1.0/255)
])
```

Figure 1. Code snapshot for Image Resize and Rescale

2.2.2. Augmentation

Augmentation is a data pre-processing technique that involves adding variations to the pictures in the dataset, such as flipping, rotation, or zooming. Increasing the dataset's diversity and strengthening the model's capacity to generalize to new images are the two objectives of augmentation. The benefits of Augmentation include increased dataset size, Improved model generalization, reduced biases and Reduced data collection efforts.

```
data_augmentation = tf.keras.Sequential([
    layers.experimental.preprocessing.RandomFlip("horizontal_and_vertical"),
    layers.experimental.preprocessing.RandomRotation(0.2),
])
```

Figure 2. Code snapshot for Data Augmentation

2.2.3. Filtering

Filtering is a common technique used in data pre-processing. To reduce noise and enhance image quality, a filter is applied to the image, such as a median or Gaussian filter. The benefits of using Filter are Noise reduction, improved image quality, reduced computation time, Standardization.

2.2.4. Splitting the Dataset

The Dataset is Spitted into three folders – train, validate and test. Images are distributed as follows: 70% are supplied to the train folder, 10% to the validate folder, and 20% to the test folder.

Table 1. Details of Plant Village dataset split for training, testing, and validation

Plant Type	Diseases Classes	Total Samples	Training Sam-ples	Test Samples	Validation Samples
Apple	Apple scab	630	441	126	63
	Apple black rot	621	434	125	62
	Apple cedar apple rust	275	192	56	27
	Apple healthy	1645	1151	330	164
Grape	Grape black rot	1180	826	236	118
	Grape esca	1383	968	277	138
	Grape healthy	423	296	85	42
Corn	Corn common rust	1192	834	239	119
	Corn northern leaf blight	985	689	198	98
	Corn healthy	1162	813	233	116
Potato	Potato early blight	1000	700	200	100
	Potato healthy	152	106	31	15
	Potato late blight	1000	700	200	100
Tomato	Tomato bacterial spot	2127	1488	427	212
	Tomato early blight	1000	700	200	100
	Tomato healthy	1591	1113	319	159
	Tomato late blight	1909	1336	383	190
	Total	18275	12787	3665	1823

2.2.5. Loss Function

The efficacy of the neural network is assessed using the loss function. By changing the model's settings during training, the objective is to reduce the loss function. The type of problem and the amount of data accessible determine the loss function to use. Cross-entropy loss is frequently used in plant disease detection because it measures the discrepancy between the actual distribution of the classes and the expected probability distribution. This project makes use of the Sparse Categorical Cross entropy loss function from the keras API of tensor flow.

```
model.compile(
    optimizer='adam',
    loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=False),
    metrics=['accuracy']
)
```

Figure 3. Code snapshot for the loss function

2.3. Model Training

In this project, the most important factor is the model training. The data gathered during the preprocess step enables the model to be trained. During training, the model learns to identify patterns and features in the images that correspond to different diseases. To improve the performance of the model, changes are made to hyper parameters such learning rate, batch size, and number of epochs. To make sure the model can generalize to new data, it is tested on a different test dataset. A lot of time must be spent training a model. For this specific project, 20 epochs are selected with a batch size of 32. Moreover, the tensor flow library is utilized, which facilitates the development and training of deep learning models. This library comes with a lot of built-in tools and methods.

Table 2. Training parameters and their values

Training parameters	Value
Epochs	20
Activation Function	ReLu
Batch Size	32
Verbose	1
Optimizer	Adam
Loss Function	Sparse Categorical Cross entropy
Metrics	Accuracy
Output Classes	17

2.3.1. Model Evaluation

To be able to assess the accuracy and functioning of the trained model, model evaluation is a critical step in the deep learning process used to detect plant diseases. With the use of multiple measures, the accuracy of model in classifying images as healthy or diseased is evaluated. Some common metrics used for model evaluation are precision, accuracy, and recall. In this project, Accuracy is used as a metric. The trained model is evaluated on the validation set to assess its performance and optimize hyper parameters. The testing set is used to assess the overall performance of the model.

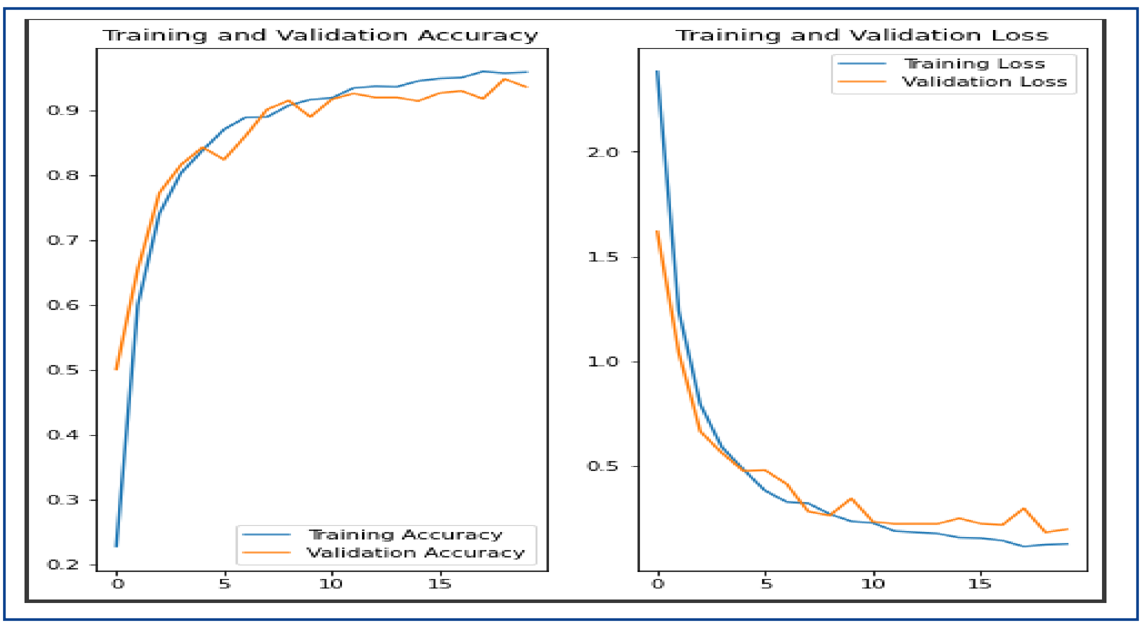

Figure 4. Accuracy and Loss Graph for Training and Validation set.

2.3.2. Model Deployment

The process of making a trained deep learning model usable in production settings is referred to as model deployment. A model must be distributed to a platform where users or other software systems can access it after it has been trained and tested.

2.3.3. Selecting a Deployment Platform

Deep learning models can be deployed on a variety of platforms, such as cloud services, containers, and edge devices. The platform to use will rely on the application's needs, including scalability, latency, and security. For this project, Docker is used as a deployment platform along with fast API as a backend service.

2.3.4. Testing the Model before Deployment

To avoid any unnecessary issues that can arise after deployment, the model must be tested to ensure that it is functioning properly. Images from the dataset are chosen, and the actual and anticipated class names are confirmed.

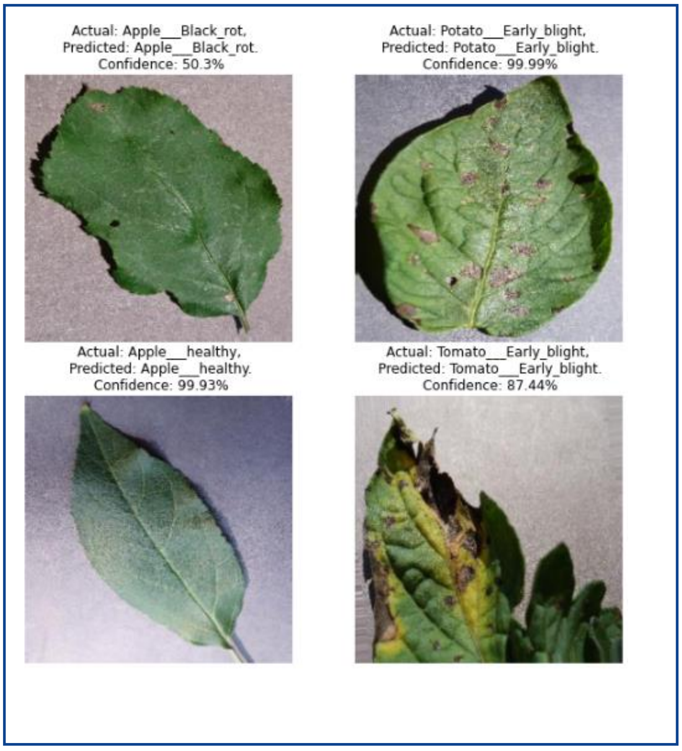

Figure 5. Snapshot of the model testing before Deployment

After the model has been tested, it is deployed using the docker deployment platform, a website built in ReactJS as the frontend, and Fast API as the backend. It uses Tensor Flow Serving with Docker [8].

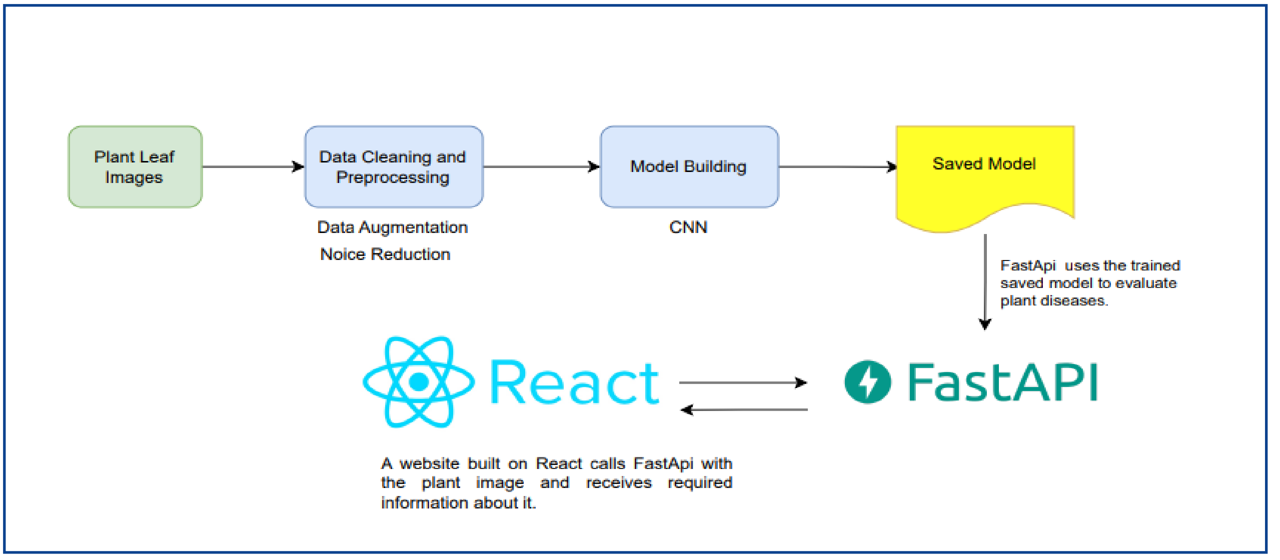

Figure 6. Block Diagram for working of project

3. Results and Discussion

After gathering the data, the next step is to analyze it utilizing deep learning and machine learning techniques. There are a variety of algorithms that might be utilised for this purpose, including deep learning, convolutional neural networks, and support vector machines. These algorithms can be trained to recognize specific patterns and characteristics in the pictures that are indicative of different plant diseases. For image processing tasks machine learning algorithms are found to be less effective. Deep learning algorithms are more suitable for these types of tasks. They have been widely used for plant disease detection because of their capacity to extract relevant features from images.

Finally, after the data has been analyzed, the algorithm can generate results indicating which plant diseases are present in the images. These results can be presented in a number of different methods, depending on the application. In this project, the result is shown on a website which is built using ReactJS. The image is delivered to the software which shows the response from the backend which is built using FastAPI and python programming language. The training, validation and testing accuracy of the model came out to be 97.12%, 94.73% and 94.13%.

Table 3. Training and validation accuracy data for last five epochs

Epochs	Training Accuracy in %	Validation Accuracy in %
16/20	96.39	92.81
17/20	96.65	96.82
18/20	96.81	95.61
19/20	96.80	96.38
20/20	97.12	94.73

Table 4. Test accuracy and loss data for test dataset

Test Loss in %	Test Accuracy in %
22.60	94.13

4. Conclusions and Future Scope

In conclusion, plant disease diagnosis using deep learning has enormous potential in revolutionizing plant disease management and increasing crop productivity. Advanced deep learning techniques such as convolutional neural networks and deep neural networks can improve the accuracy of plant disease diagnosis, and integration with IoT, cloud-based solutions, and mobile applications can provide real-time monitoring of plant health and disease outbreaks. However, there are also limitations, such as limited data availability, limited access to technology, and high equipment costs, which need to be addressed. Collaboration between researchers, farmers, and other stakeholders can help address the limitations and ensure the successful adoption of deep learning-based solutions for plant disease diagnosis. With continued development and implementation, utilizing deep learning to identify plant diseases could potentially change the agricultural industry and help address food security challenges. Deep learning-based plant disease diagnosis is a potent instrument that has the potential to transform agriculture by providing prompt and accurate diagnosis of plant diseases, lowering the likelihood of yield loss, and increasing crop productivity. However, additional study is required to increase the precision of these models, make them more accessible to farmers, and integrate them into existing farming practices.

References

1. Mohanty SP, Hughes DP, Salathé M. Using deep learning for image-based plant disease detection. Front Plant Sci [Internet]. 2016;7:1419. Available from: http://dx.doi.org/10.3389/fpls.2016.01419
2. New standards to curb the global spread of plant pests and diseases [Internet]. Fao.org. [cited 2023 May 8]. Available from: https://www.fao.org/news/story/en/item/1187738/icode/
3. Andrew, Eunice J, Popescu DE, Chowdary MK, Hemanth J. Deep learning-based leaf disease detection in crops using images for agricultural applications. Agronomy (Basel) [Internet]. 2022;12(10):2395. Available from: http://dx.doi.org/10.3390/agronomy12102395
4. PlantVillage Dataset. PlantVillage Dataset | Kaggle.com. [cited 2023 May 7]. Available from: https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset.
5. Tan L, Lu J, Jiang H. Tomato leaf diseases classification based on leaf images: A comparison between classical machine learning and deep learning methods. AgriEngineering [Internet]. 2021;3(3):542–58. Available from: http://dx.doi.org/10.3390/agriengineering3030035
6. Wang L, Lee C-Y, Tu Z, Lazebnik S. Training deeper convolutional networks with deep supervision [Internet]. arXiv [cs.CV]. 2015. Available from: http://arxiv.org/abs/1505.02496
7. Kartikeyan P, Shrivastava G. Review on emerging trends in detection of plant diseases using image processing with machine learning. Int J Comput Appl [Internet]. 2021;174(11):39–48. Available from: http://dx.doi.org/10.5120/ijca2021920990
8. TensorFlow serving with docker [Internet]. TensorFlow. [cited 2023 May 8]. Available from: https://www.tensorflow.org/tfx/serving/docker