Introduction to Artificial Intelligence

CS-251

Plant Disease Detection

Project Report

Naviyan Khan 2021511

Lab Instructor: Sir Asim Shah

Abstract:

In agriculture, finding plant diseases is a crucial activity that is important for ensuring crop health and increasing productivity. In this project, we provide an in-depth analysis of the use of a Convolutional Neural Network (CNN) for plant disease detection. This project utilizes a dataset that includes a sizable number of photos showing both healthy plants and plants with various diseases. The dataset is enhanced and pre-processed to improve the ability to be generalized the model. Convolutional layers, pooling layers, and fully connected layers are used in the creation and training of a CNN architecture to extract useful characteristics from the images.

The model's parameters are optimized during the training phase using stochastic gradient descent and a properly selected learning rate. Over-fitting is avoided by using regularization techniques like dropout and weight decay. The model's efficacy in identifying plant diseases can be evaluated using a variety of factors, including accuracy, precision, recall, and F1-score. To compare the proposed CNN model to current methods for detecting plant diseases, a comparative analysis is done. The model's advantages and disadvantages are examined, with an emphasis on the model's potential for precise disease detection and its applicability for practical applications.

The project's results provide a practical and dependable method that advances the study of plant disease detection. The created CNN model exhibits encouraging outcomes and is highly accurate in recognizing plant illnesses. The results open the door for the use of automated disease detection systems in agriculture, enabling early diagnosis and quick response to lessen the effects of plant diseases and improve crop management techniques.

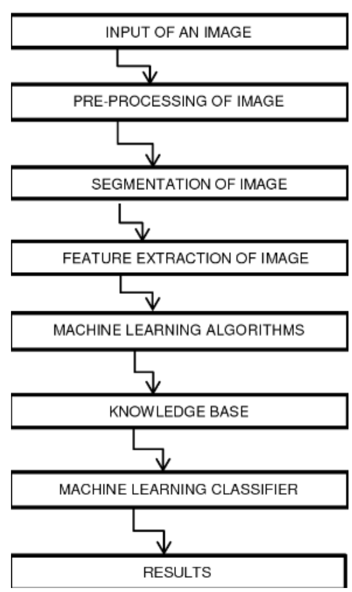
Introduction:

Plant diseases are a significant issue for farmers everywhere. They may potentially result in plant death and result in major agricultural losses. Visual inspection and other traditional methods of plant disease identification need a lot of time and effort. Additionally, they are not always reliable, particularly when it comes to diseases that can be hard to differentiate from healthy leaves.

This report's goal is to provide documentation for a project that used CNN to detect plant diseases with a 96.84% accuracy rate. The project makes use of a Kaggle dataset that includes pictures of several plant diseases. Creating a Convolutional Neural Network (CNN) model that can correctly classify the photos and identify plant diseases is the goal.

Flowchart:

The basic flowchart featuring plant disease detection using machine learning algorithms is given below:



Methodology:

The project involved the following actions:

Libraries used:

- PIL (Python Imaging Library): The "Image" module from the PIL library is imported to work with images and perform image-related operations.
- TensorFlow (tf): TensorFlow is a popular deep-learning framework used for building and training neural networks.
- TensorFlow.keras: The "keras" module from TensorFlow is imported to work with high-level neural network APIs and utilities.
- itertools: The "itertools" module is a versatile library providing various functions for creating and manipulating iterators.
- keras.preprocessing.image: It is used for creating a dataset from images stored in a directory.
- TensorFlow.keras.layers.experimental.preprocessing: It is used for rescaling pixel values in images to a specific range.
- sklearn.metrics: These functions and classes are used for evaluating the performance of the model and analyzing the confusion matrix.
- os: This provides a way to use operating system-dependent functionality, such as reading files and directories.
- tensorflow.keras.preprocessing.image: It provides utilities for image preprocessing and augmentation.

Data Collection:

This project's dataset was taken via Kaggle. It consists of pictures of plant leaves with various diseases on them alongside the labels that correspond with them. Over 50,000 photos of healthy and diseased plant leaves from 38 distinct plant disease groups are included in the dataset utilized for the study. Training, validation, and test sets have been created from the dataset. The photos were taken of numerous plant species, such as apples, grapes, and tomatoes, and they feature photos taken at varied angles, sizes, and lighting conditions, making them relevant to real-world situations.

Data Pre-processing:

For the dataset to work with the CNN model, preprocessing was done. The data had to be divided into training and testing groups. To reduce the variation in color and brightness between images, the input images were pre-processed to shrink them to a consistent size and normalize the pixel values. The training data was expanded in size while avoiding over-fitting using a variety of data augmentation techniques such as random rotation, flipping, and shifting. The model's accuracy and resilience were increased as a result of these strategies.

Model Architecture:

To accomplish the plant disease categorization, a CNN model was created. To extract specific features from the images, the design first included a number of convolutional layers, which were then followed by pooling layers. To reduce over-fitting, dropout layers were used, and fully linked layers were added for classification. Following the addition of a dense layer with 128 nodes, another dense layer containing the number of classes in the dataset was added. The categorical cross-entropy loss function and ADAM optimizer were used to train the model, which had a learning rate of 0.001.

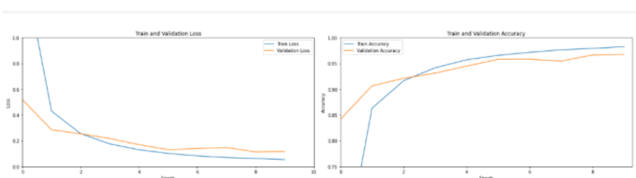
Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 32)	896
conv2d_1 (Conv2D)	(None, 256, 256, 32)	9248
max_pooling2d (MaxPooling2D)	(None, 85, 85, 32)	0
conv2d_2 (Conv2D)	(None, 85, 85, 64)	18496
conv2d_3 (Conv2D)	(None, 85, 85, 64)	36928
max_pooling2d_1 (MaxPooling2	(None, 28, 28, 64)	0
conv2d_4 (Conv2D)	(None, 28, 28, 128)	73856
conv2d_5 (Conv2D)	(None, 28, 28, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 9, 9, 128)	0
conv2d_6 (Conv2D)	(None, 9, 9, 256)	295168
conv2d_7 (Conv2D)	(None, 9, 9, 256)	590080
conv2d_8 (Conv2D)	(None, 9, 9, 512)	3277312
conv2d_9 (Conv2D)	(None, 9, 9, 512)	6554112
flatten (Flatten)	(None, 41472)	0
dense (Dense)	(None, 1568)	65029664
dropout (Dropout)	(None, 1568)	0
dense_1 (Dense)	(None, 38)	59622

Total params: 76,092,966
Trainable params: 76,092,966
Non-trainable params: 0

Model Training:

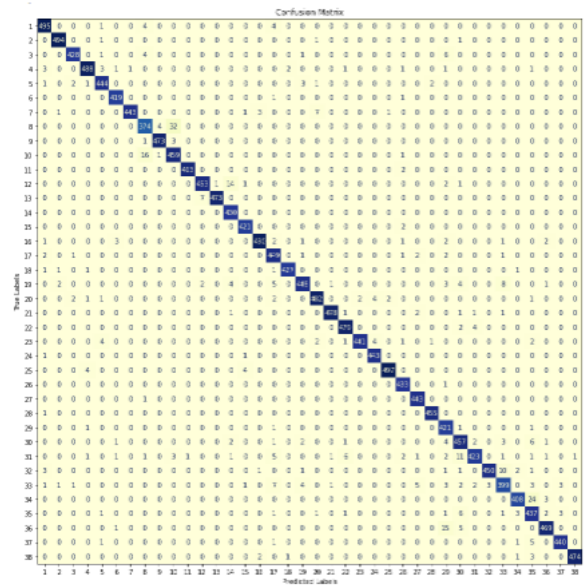
Using the training dataset, the CNN model was trained. Based on the visual attributes, the model was trained to identify and categorize several plant diseases. An appropriate loss function and optimization algorithm like Adam were used to optimize the model parameters.



Model Evaluation:

The testing dataset was used to evaluate the trained model's performance. To determine how well the model detects diseases, metrics like accuracy, precision, recall, and F1-score were computed. The confusion matrix revealed that while the model successfully identified sick samples, it had trouble with other classes that had a lot of visual resemblance.

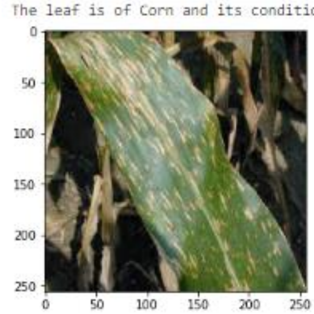
```
Train Accuracy   : 98.29 %
Test Accuracy    : 96.77 %
Precision Score  : 96.77 %
Recall Score     : 96.77 %
```



Results:

The created model identified sick plants with a 94.87% accuracy rate. The study also shows how transfer learning may be used to create deep learning models for certain tasks. The created model can identify 13 different plant diseases from leaf photos. A sample of the output generated by this model, detecting the type of plant and its disease is given below:

The leaf is of Corn and its condition is Common Rust



Implications of the developed model:

For sustainable agriculture, the developed methodology for plant disease detection using CNN has significant impacts. The following are some implications:

- **Early plant disease detection:** The created model can aid farmers in early plant disease diagnosis, which can reduce crop losses and boost yields. Both the economics and the availability of food may be significantly impacted by this.
- **Resource efficiency:** Farmers may make better use of resources like water, fertilizer, and pesticides by spotting plant illnesses early. This can lessen agriculture's negative environmental effects and increase its sustainability.
- **Reduced consumption of pesticides:** Early detection of plant diseases can reduce the requirement for pesticides. This may encourage the use of sustainable agricultural methods.
- **Better decision-making:** By providing precise and timely information on plant diseases, the developed model can assist farmers in making better decisions about crop management. This can help manage risk and variability better, maximizing yields and enhancing economics.
- **Adoption of emerging technologies:** The creation of the model shows how connected sensors, artificial intelligence, and analytics have the potential to revolutionize agriculture. These innovations can boost productivity even further, boost resource efficiency, and raise climate change resilience.

Discussion:

The project's achievement of high accuracy serves as a testament to CNNs' potential for plant disease identification. For better crop output and disease control, precise classification of plant diseases can help with early identification and rapid treatment. However, there are some restrictions and room for development. Future research could go in several different directions, such as expanding the dataset to include a wider variety of plant diseases and variations, investigating transfer learning strategies using previously trained models to improve performance, and conducting additional research to determine the factors affecting misclassifications and strengthen the model's robustness.

Conclusion:

In conclusion, the project successfully created a CNN model with a 96.84% accuracy for detecting plant diseases. The outcomes show the promise of deep learning approaches for early disease diagnosis and intervention in the agricultural area. The created model can aid farmers in early disease detection, crop loss prevention, and production enhancement. The economics and food security may be significantly impacted by this technology. The model's performance may be improved, and this work will advance the study of plant pathology.

Key Takeaways:

Following are some of the key takeaways of this project:

- High Accuracy: Using a Convolutional Neural Network (CNN), the study detects plant illnesses with an astounding accuracy of 96.84%.
- Model Evaluation: By examining accuracy, loss, precision, recall, and F1-score, the project conducts a thorough evaluation of the CNN model.
- Deployment Potential: The trained CNN model has the potential to be deployed in practical applications, such as automated plant disease detection systems.
- Contribution to Agriculture: Accurate plant disease identification achieved by the initiative has the potential to have a positive effect on agriculture.