Deep Learning Based Recognition of Plant Diseases

Swetha Parthiban, Sneha Moorthy, Sribalaji Sabanayagam, Shobana Shanmugasundaram, Athishwaran Naganathan, Mohan Annamalai, and Sabitha Balasubramanian

Abstract Every country's primary requirement is agricultural products. Infected plants have an impact on the agricultural production and economic resources of a country. Early detection of plant diseases minimizes the risk of crop loss by allowing farmers to adopt curative and preventative actions to avoid further damage. Citrus is a large plant that is primarily grown in tropical areas of the world because of its high vitamin C and other key nutrient content. Citrus plant diseases play a vital role in causing a great financial loss to the farmers and affect the economy of the country. The tomato ranks among the most economically important vegetable crops, with a global production rate of around 200 million tonnes. But due to diseases occurring in the tomato plants, the farmers are facing a great loss. To prevent these losses and provide an immediate cure, we create a model to detect plant diseases. To get a better performance model, the data augmentation process is used, and we got 18,392 images. In our model, we used a convolutional neural network (CNN) to classify the citrus and tomato leaf diseases. The diseases are Mancha Graxa, Citrus Canker, Tomato Early Blight, Tomato Septoria Leafspot, Tomato Spider Mite Two-spotted spider mite (Tetranychus urticae), Tomato mosaic virus (Tobamovirus), Tomato yellow leaf curl virus (Begomovirus), Tomato Target Spot (Corynespora cassiicola). Convolution Neural Network (CNN) is stacked by multiple convolutions and pooling layers to detect plant leaf diseases. We train a model with augmented images. After the model gets trained, it is thoroughly tested to ensure that the results are accurate. Thereby we were able to achieve high validation accuracy with a rate of (97.02%) for this dataset used. As the results showed, applying the deep CNN model to identify diseases could greatly impact the accurate identification of diseases, and could even prove useful in detecting diseases in real-time agricultural systems.

S. Parthiban (✉) · S. Moorthy · S. Sabanayagam · S. Shanmugasundaram · A. Naganathan · M. Annamalai · S. Balasubramanian
Kumaraguru College of Technology, Coimbatore, India
e-mail: swetha.20ei@kct.ac.in

S. Balasubramanian
e-mail: Sabitha.b.mce@kct.ac.in

1 Introduction

In recent years, digital image processing, image analysis, and machine vision have
evolved dramatically, and they have become a key component of artificial intelligence
and the link between grounded theory and applied technology. These technologies are
widely utilized in industry and medicine, but they are only seldom used in agriculture
and natural ecosystems. In the domain of image processing and Deep learning, as
well as its application in numerous fields of engineering, significant advances have
been made. Agriculture research aims to boost food production and quality while also
reducing costs and increasing profitability. Fruits are high in vitamins and minerals, as
well as fibre and other non-nutrient compounds. Fruit from citrus plants contains high
levels of vitamin C, which has multiple health benefits and is also used as raw material
in many agricultural industries. Various diseases damage citrus fruit plants, including
citrus canker and Mancha graxa. Infectious citrus canker causes yellow halo lesions
as well as yellow spots on leaves and fruits. Cancer bacteria spread easily and quickly
through clothing and infected components to air currents, plants, birds, and humans.
Greasy spot is a disease that is common in citrus cultivars produced in tropical
and semitropical climates. Mycosphaerella citri is the cause of the sickness. On the
underside of mature citrus leaves, symptoms show as yellow to dark brown to black
lesions. Infected cells fail to generate chlorophyll, resulting in these yellow (chlorotic)
blotches. On lemons and grapefruit, lesions are more yellowish and widespread,
whereas on tangerines, lesions are more elevated and darker. Tomato Early Blight
disease is caused by the fungi Alternaria linariae, Lesion size can be substantial,
usually surrounding the entire fruit. For infected fruit, concentric rings can also be
seen. Tomato Septoria leafspot is caused by the fungus Septoria lycopersici, older
leaves reveal a large number of small, circular spots with dark borders surrounding
a beige center. Tiny black flecks must be spore-producing bodies located within
the centers of the spots. In the agricultural sector, the Tomato Spider mites Two-
spotted spider mite is a widespread pest that affects nearly all plants (more than
1000 species). Tomato Target Spot is caused by Corynespora cassiicola and this
disease causes severe foliar damage to greenhouse and field tomatoes. The tomato
yellow leaf curl virus is part of the plant family Geminiviridae and belongs to the
genus Begomovirus and throughout the tropical and subtropical regions, the disease
causes severe economic losses. The tomato mosaic virus is a plant pathogenic virus
and a member of the class tobamovirus. In the family tobamoviridae, it belongs
to the genus Tobamovirus, and this disease is easily spread by contact, cultural
practices, or contaminated seeds. It is extremely important to identify plant diseases
early in the field in order to manage their detection and spread. The conventional
approach to disease identification relies on agricultural extension organizations for
support, but this approach is expensive and limited in countries with low resources

and infrastructure. In this case, Deep learning helps us to sought out this problem because deep learning has the advantage of requiring minimal domain expertise because no feature engineering is required. An automatic plant disease detection system based on machine learning algorithms detects diseased plants and the type of disease they have by using computer vision. Convolution neural network (CNN) can be used as a deep learning network for images to do this. CNN is used to extract visual properties such as horizontal and vertical edges, RGB values, and so on. For visual feature extraction, CNN is the finest deep learning neural network. By giving many photos of diseased plants, a CNN-based network may be trained to detect disease in plants, and the trained model can then be used to forecast disease in plants using photographs of plant leaves in the future.

2 Literature Review

Chohan et al. [1] suggested a deep learning model for detecting plant pathogens. They apply augmentation to increase the data before training the model with a convolutional neural network. They obtain an accuracy of over 98.3%. Senthilkumar et al. [2] presented a method for identifying and classifying citrus plant diseases it constitutes of 4 processes that is pre-processing, segmentation, feature extraction, and classification. Otsu is used for segmentation. Inception ResNet is used for feature extraction. For classification, they used random forests. They got 98.91% accuracy. Zeng et al. [3] developed a deep learning model to detect citrus plant diseases with help of inception_v3 architecture in addition to that they used GANs based data augmentation to increase the performance. Huang et al. [4] proposed two models one to detect plant disease and other is to plant classification. They obtain accuracy around 98% for plant classification and 87% for disease identification. Kukreja et al. [5] proposed a deep learning model for the detection of citrus fruit diseases using the process of data augmentation and pre-processing. They attain accuracy around 89.1%. Mohanty et al. [6] trained a deep convolutional neural network to 12 to identify 14 crop species and 26 diseases with 54,306 images and paved the way for smartphone-assisted disease diagnosis. They obtain accuracy around 99.35%. Song et al. [7] proposed a model to detect citrus plant diseases using YOLO (you only look once) algorithm. YOLO can detect and circle around a disease in real time on an image or video. Jasim et al. [8] proposed a robust methodology to detect and classify the diseases. They obtain accuracy around 98%. Arya et al. [9] have analyzed the various research works and reviewed the architecture used for the detection of diseases in plant leaves. Alruwaili et al. [10] used the AlexNet architecture and it was customized in order to detect disease in an efficient manner. Then they use new augmentation method. To overcome the overfitting, they used "sgdm" method.

3 Scope

The scope of our project is to maximize productivity and ensure agricultural sustainability. This model will serve as a benefit for the farmers who have cultivated citrus and tomato plants. Farmers get benefitted from this model with high yield and minimize the economic loss and it gives a positive impact on the farmers.

4 Methodology

4.1 Procedure

1. Datasets

In computing, a dataset is a collection of related, discrete pieces of related information that can be gotten at exclusively or mix-and-matched, or that can be tracked in aggregate. Figure 1 shows the diseased leaf images.

2. Data Augmentation

Adding augmentation data to existing data sets or creating new synthetic data sets from existing data is a process of increasing the amount of data in data analysis. This function serves as a regularizer and reduces overfitting when training a model. It has a lot to do with data analysis oversampling. Through the process of data augmentation, we have increased our original dataset to double the size. Consequently, we were able to create adequate datasets and train our model to achieve maximum accuracy.

Fig. 1 Leaf disease image

Table 1 I: CNN training parameters

Parameter	Value
Epochs	25
Batch size	32
Activation in middle layer	ReLU
Activation in final layer	SoftMax

3. Environment

Execution of code needs an access to a system with GPU. For this, we used an online platform called Google colab as these notebooks have GPU access. To detect the citrus and tomato diseases, we use Convolutional Neural Network Algorithm.

4. Convolutional Neural Network Algorithm

The Input layer, Middle layers, and Output layer make up a Convolutional Neural Network, or CNN. The input layer is responsible for accepting features such as input, or images. The number of middle layer nodes is determined by the application. A result is created by the output layer

Convolutional Layer-Convolution is performed on the pixel values in conjunction with the kernel matrix. The kernel matrix is dragged across the pixel matrix to calculate the result.

Max Pooling Layer-Filter maps are generated using this layer to minimize their size. This reduces the likelihood of overfitting issues.

Rectified Linear Units (ReLU) Activation Function-With the Rectified Linear Units (ReLU) Activation Function, all negative values are replaced by zero (0) while all positive values are maintained.

Fully Connected Layer-A node from the previous hidden layer is connected to a node from the next hidden layer in this layer, which is also known as Dense Nodes. Each layer will be connected by use of edge connections among the neurons. Here Table 1 shows the training parameters of the CNN model.

4.2 Library

Keras is a popular deep learning library that runs on top of other libraries such as tensor flow. Its minimal Python structure facilitates learning and quick writing of deep learning models.

Algorithm

STEP 1: Importing the libraries and initializing the data path.

STEP 2: Importing the datasets.

STEP 3: preprocessing—write a function to resize the dataset images to (224,224) so that they can be fit for training. Also visualizing our data and dividing it into training and testing categories.

STEP 4: Defining train and test data into the model. The train dataset has 12,723 images belonging to 8 classes and the test dataset has 5669 images belonging to 8 classes.

STEP 5: we use keras library to develop a plant disease detection model. Because it has good support for functions that enable users to deploy models fast. Keras supports building neural network models with fewer lines of code.

STEP 6: we create a sequential model for this plant disease classification. we develop the 1st block of 2D convolutional layers with 32 filters of 5*5 kernel and using ReLU (rectified linear unit) activation function then we perform max pooling operation in the layers. Secondly, we used 32 filters of the 3*3 kernel with ReLU (rectified linear units) activation function in the layers and then we performed maximum pooling. In the 3rd block of 2D convolutional layers, we have 64 filters of 3*3 kernels, one of which is active in a ReLU (rectified linear unit) activation function, and the other operates in a Max Pooling mode.

The next layer is a flattened layer. We used 25% (0.25) dropout function in the dense layer.

STEP 7: Here, we declare all the hyper parameters required for our plant disease classification such as epochs, steps, batch size, learning rate etc. Fit our training data to our model and set the batch size to 32, which will accept 398 values at a time until all parameters are met. The number of epochs here refers to the number of times it will be processed.

STEP 8: Training the Model-The model has been developed, and it should be compiled using the Adam optimizer, which is one of the most familiar optimizers.

STEP 9: We take 25 epochs to train our model. More epochs increase the accuracy and decrease the loss. we can see here that accuracy in each step increase. we train our model 25 times. When test images train with lots of train images then accuracy will increase, and our model predicts more accuracy. More accuracy means our model works properly and gives desired output. Figure 2 gives the summary of CNN model architecture.

STEP 10: Detection of the disease when the input is given.

STEP 11: Computing Loss Calculation Based on Training and Test Results.

STEP 12: Accuracy Calculation for Training and Test Results-Matplotlib was used to show the accuracy plot on the training and test sets. A simple graph analysis can clearly reveal the difference in training and test set accuracy.

Accuracy = Number of accurately predicted records/Total number of records.

```
Model: "sequential"
[ ]
  Layer (type)                    Output Shape              Param #
  =================================================================
  conv2d_1 (Conv2D)              (None, 220, 220, 32)       2432

  max_pooling2d_1 (MaxPooling2   (None, 73, 73, 32)         0

  conv2d_2 (Conv2D)              (None, 71, 71, 32)         9248

  max_pooling2d_2 (MaxPooling2   (None, 35, 35, 32)         0

  conv2d_3 (Conv2D)              (None, 33, 33, 64)         18496

  max_pooling2d_3 (MaxPooling2   (None, 16, 16, 64)         0

  flatten_1 (Flatten)            (None, 16384)              0

  dense (Dense)                  (None, 512)                8389120

  dropout (Dropout)              (None, 512)                0

  dense_1 (Dense)                (None, 128)                65664

  dense_2 (Dense)                (None, 8)                  1032
  =================================================================
  Total params: 8,485,992
  Trainable params: 8,485,992
  Non-trainable params: 0
```

Fig. 2 Convolutional layers

The result is 0.9702, which means the model is 97.02% accurate in making a correct prediction. The flow of the project is pictured in Fig. 3.

5 Results and Findings

This study emphasizes the need of early diagnosis of plant diseases. Deep Learning was used to create this model. The accuracy of this model was tested using 30% (5669) of photos. The photographs in this gallery are from eight distinct classes. For testing, 30% of each class was chosen at random. The correctness of the testing dataset is greater than 95%. Model properly Classified 5669 photos out of 18,392 images. On the testing dataset, our model generated the Training and Validation accuracy curve as shown in Figs. 4 and 5.
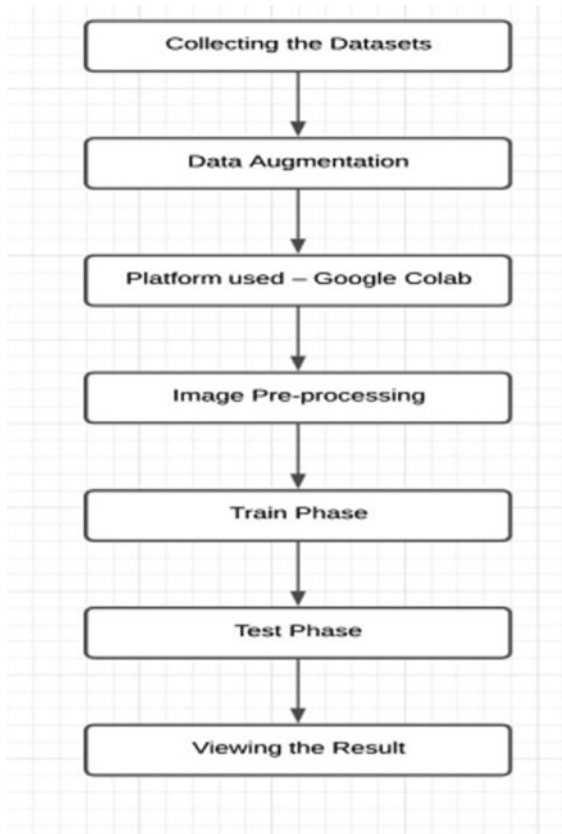
FLOWCHART.

Fig. 3 Work flow

6 Discussion

To identify citrus and tomato illness with high efficiency, this method primarily uses Convolutional Neural Networks. When verification performance and model storage are constrained, the CNN model can be useful. An important component for diagnosing and identifying crop diseases and insect pests is the availability of rich and diverse databases, and a large dataset can decrease the error rate in the recognition process. The power of deep learning tools can only be fully unleashed by accumulating. In addition, we intend to acquire ever larger databases in order to improve the model's generalization capabilities. As the dataset collected on citrus diseases is being analyzed, the backgrounds can be removed, allowing better disease images to be used as training models for improving the accuracy of disease identification. It is possible
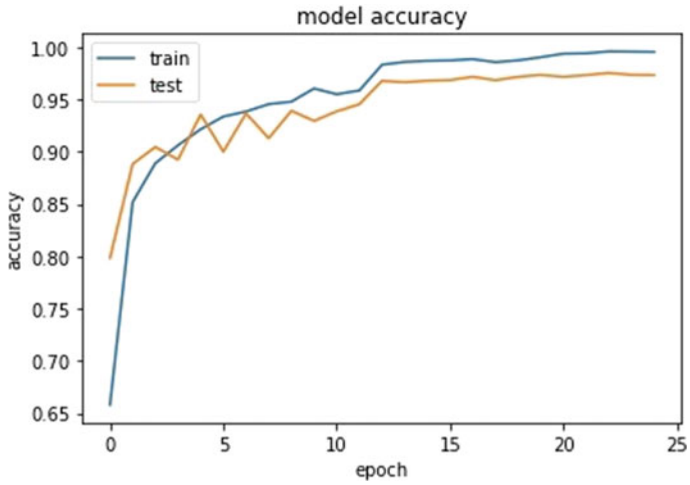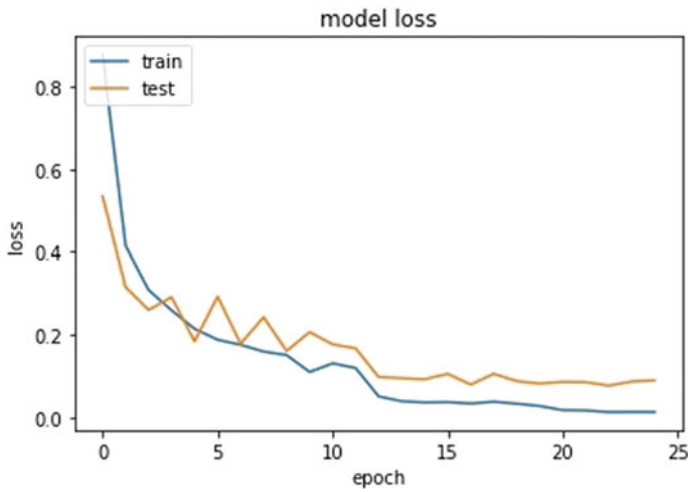
Fig. 4 Model accuracy

Fig. 5 Model loss

to attain identification accuracy by optimizing the convolution network algorithm. Optimisation of the convolution operation can ensure more accurate identification of data obtained from larger datasets.

7   Conclusion

Using a form of modern approaches, plant diseases are detected and classified using sick leaves. Commercial solutions to spot these diseases aren't available at the current time. In our work, we used CNN models for the detection of plant diseases using diseased-leaf images of citrus and Lycopersicon esculentum. The quality dataset with 18,392 images was accustomed to train and test the model (after data augmentation). There are 8 different classes of plant diseases namely Mancha Graxa, Citrus Canker, Tomato Early Blight, Tomato Septoria Leafspot, Tomato mite Two-spotted mite (Tetranychus urticae), Tomato mosaic virus (Tobamovirus), Tomato yellow leaf curl virus (Begomovirus), Tomato Target Spot (Corynespora cassiicola).

After splitting the dataset into 70–30 (70% data for training, 30% data for testing), 97.02% accuracy was achieved in our model. On average, it took 25 and 25 s/epoch, respectively, on coloured images. With relevance to both accuracy and loss, the implemented deep-learning model reduces the deviation from the best deep learning model. The desired time to coach the model was much but that of other machine-learning approaches.

8   Future Scope

The present model is modified to measure the detection model and it is designed as an internet application which helps farmers to make a decision on the particular quantity for pesticide applications to scale back the price and environmental pollution and increase the productivity. Moreover, the system predicts the kind of plant and kind of disease so scientists can recommend specific pesticides or fungicides to be used, thus providing useful suggestions for future research. Detecting a disease during this crop is predicated solely on the leaf of the plant. Roots, stems, and branches from the crop are easier to handle if they will be included yet which increases the detection accuracy. It'll also show the disease name if the model receives input aside from leaf images. And also had a concept to incorporate another crop diseases to extend the predictability.

9   Recommendation

There are countless new plant diseases reported annually or an old disease that has reappeared and is now affecting crops previously unaffected. But the notice towards the detection of plant diseases in an early stage isn't for sure. The fundamental thought when the plant gets affected is to use fertilizers and a few chemicals utilizers but while taking concern about the health of the plant one of all the most effective ways is early detection. In the event of controversy during the growing process, ask

questions or keep an eye fixed on the matter. By following the recommendations, the impact of plant diseases is greatly reduced, leading to a decrease in shrinkage and increased profit.

References

1. Chohan M, Khan A, Chohan R, Hassan S, Mahar M (2020) Plant disease detection using deep learning. Int J Recent Technol Eng 9(1):909–914
2. Senthilkumar C, Kamarasan M, A novel citrus disease detection and classification using deep learning based inception resnet V2 model
3. Zeng Q, Ma X, Cheng B, Zhou E, Pang W (2020) GANs-data augmentation for citrus disease severity detection using deep learning. IEEE Access 8:172882–172891
4. Huang S, Liu W, Qi F, Yang K (2019) Development and validation of a deep learning algorithm for the Recognition of plant disease. In: 2019 IEEE 21st international conference on high performance computing and communications; IEEE 17th international conference on smart city; IEEE 5th international conference on data science and systems (HPCC/Smart City/DSS). IEEE, pp 1951–1957
5. Kukreja V, Dhiman P (2020) A deep neural network based disease detection scheme for citrus fruits. In: 2020 international conference on smart electronics and communication (ICOSEC). IEEE, pp 97–101
6. Mohanty SP, Hughes DP, Marcel S (2016) Using deep learning for image-based plant disease detection. Front Plant Sci 7:1419
7. Song C, Wang C, Yang Y (2020) Automation detection and image recognition of precision agriculture for citrus diseases. In: 2020 IEEE Eurasia conference on IOT, communication and engineering (ECICE). IEEE, vol 23, pp 187–190
8. Jasim MA, AL-Tuwaijari JM (2020) Plant leaf diseases detection and classification using image processing and deep learning techniques. In: 2020 international conference on computer science and software engineering (CSASE). IEEE, pp 259–265
9. Arya S, Singh R (2019) An analysis of deep learning techniques for plant leaf disease detection. Int J Comput Sci Inf Secur (IJCSIS) 17(7)
10. Alruwaili M, Abd El-Ghany S, Shehab A (2019) An enhanced plant disease classifier model based on deep learning techniques. Int J Eng Adv Technol 9(1):7159–7164