Plant Disease Detection Using CNN

P. Silpa Chaitanya[1], K.Harshini[2], K.Moni Priyanka[3], K.Pranavika Sri[4], D.Pavani[5]

Vignan's Nirula Institute of Technology and Science for Women

Corresponding author: P.Silpa Chaitanya Email: silpam86@gmail.com

Plant disease identification is the study based upon the fact of patterns. Plant diseases make production more difficult to produce. To avoid losses in agricultural productivity and quantity, disease detection is used. Deep learning can help to avoid the problems of manually choosing disease spot features, raise the accuracy of plant disease extracting features, and increase speed scientific and innovation transition. The combination of increased smartphone usage and deep learning-enabled advancements in computer vision has paved the path for smartphone-assisted virus diagnosis. We used images to train the data and perform the CNN algorithm. One of the key concerns that determines the deficit of harvest production and agricultural production is the detection and discovery of plant diseases. Plant infection investigations are the examination of any visible focus in any part of the plant that aids us in distinguishing any two plants, in particular any spots or shading conceals. One of the most important factors in the horticultural turn of events is the plant's ability to be maintained. It's exceedingly difficult to get the appropriate identifiable proof of plant illnesses. The recognizable proof of illness needs considerable effort and experience, as well as extensive understanding of plants and the research into the diseases identification.

Keywords: Deep Learning, Plant Disease, Segmentation, Classification, Feature Extraction, Image Processing; Convolutional Neural Network

2023. In Satyasai Jagannath Nanda & Rajendra Prasad Yadav (eds.), *Data Science and Intelligent Computing Techniques*, 47–54. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.56155/978-81-955020-2-8-5

1 Introduction

India has the most agriculturally based economy on the earth. Approximately 75% of the population is directly or indirectly dependent on agriculture. Disease-free crop production is critical for the growth of the country's agricultural sector. Historically, agricultural organizations such as local plant clinics have aided in disease detection. Diseases are very common in plants due to a variety of factors such as fertilizers, cultural practices, environmental conditions, and so on. These diseases harm agricultural fields and have a negative impact on the economy. In agriculture, it's critical to monitor the health of the crops on the farm. Experts in the field manually identify these diseases. This entails a significant amount of work as well as it takes lengthy processing time. The disease symptoms can be identified on the stem, leaves and fruits of the plant.

Fig. 1: Sample Leaf Image

Machine learning (ML) and artificial intelligence (AI) are both branches of deep learning (AI). To learn new knowledge, deep learning creates more complicated hierarchical models like mimicry. Deep learning algorithms, often the structure of the human brain, are used as the inspiration for neural networks. These neural networks are made up of interconnected network switches that learn to detect patterns over time. Deep Learning is concerned with improving the machine's ability to learn new things. Deep learning is used by the data scientist to transform raw data into algorithms. The system then examines the data without regard for rules or features. Then the system makes predictions, they are validated against a separate set of data. The term "deep learning" refers to the accumulation of many layers of the neural network over time, with performance increasing as the network becomes deeper. In each level of the network, the input data is processed in a specific way, which then informs the next layer. As a result, the one layer's output becomes the next layer's input. Deep learning is extremely powerful and adaptable, which is a part of machine learning. Among the different Networking architectures technologies used in Deep learning, Convolutional Neural Networks are based on image acknowledgement. In the Convolutional Neural Network it has one or more convolutional layers which are used for image processing, classification, segmentation, and auto correlated data. It's a filter that slides over the input. CNN classifiers are trained to detect diseases in individual plants. Convolutional neural networks are used to extract feature maps and image classification. Unlike traditional machine learning techniques, CNN optimizes the weights of the hidden layer and the filter parameters. Create features are appropriate for solving the categorization challenge to reduce the classification error, the network's parameters are tuned using gradient descent and back- propagation techniques.

2 Literature survey

The spread of diseases on the plants increases rapidly, it is essential to provide the solution for the affected plants and identify the diseases. The modern technology is improved to recognize the diseases on the plant and gives good accuracy. CNN is used by the image objects. CNNs have gained popularity in recent years, with DL being one of the most popular architectures. At several different convolutional levels, DL models can learn valuable properties from the input images. DL was able to tackle complicated issues successfully and quickly with a high accuracy and a low failure rate. CNN model has many layers from layer 1 to layer 16. Due to Batch normalization and dropout, Regularization is attained [1]. The batch normalization technique is used, it will transform the high values into ranges from 0 to 1. Dataset is collected from the Plant village open source. Python is used for the

augmentation process in this project. The transformation of the image is done by using the Image Code Generator class. The Image Code Generator allows the user to rotate images randomly to any degree from 0 to 360. The different cases provide the true and false conditions which are identified and classified.

The authors of [2] used images of tomato leaves to identify diseases. They classified diseased segments using geometrical and framework matching features, as well as a SVM classifier with polynomial kernel. S.Kaur et al. [3] used color and textural parameters to identify three distinct soybean illnesses. P Babu combined a neural network for identifying plant leaves and a neural network for back propagation and illnesses in [4].

Plant Disease Identification for Smart Agriculture Using bModel-based Statistical Attributes was proposed by Chit Su Hlaing and SaiMaung MaungZaw. The merits of the GP conveyance methodology for SIFT labels have been demonstrated, as well as how it may be used efficiently in plant sickness order. In their study, Malvika Ranjan et al. by capturing photos of damaged leaves, presented a technique for detecting illnesses in plants "Leaf Disease Detection and Classification Using Artificial Neural Network". By carefully selecting feature values, An Artificial Neural Network (ANN) is given training to distinguish between damaged and healthy plant samples. The accuracy of the ANN model is 80 percent.

According to the article, "textural properties of plant leaves are used to detect sick regions and classify plant leaf diseases". [5]The disease detection technique, as per S. Arivazhagan consists of four main steps: The input RGB image will be first provided a color transformation architecture, and afterwards green pixels are removed using a certain threshold value discovered and uninvolved pixels are left, that is then preceded by a segmentation stage and texture statistics to obtain beneficial segments. Finally, the characteristics retrieved for disease classification are fed into a classifier.

In their article "Applying image processing technique to detect plant diseases," Using an Artificial Neural Network and a variety of image processing techniques, Kulkarni et al. created a system for accurate and fast plant disease detection. Because it uses ANN classifiers for classification and only a Gabor filter for extracting features, the proposed method provides more accurate results, with a recognition rate as high as 91 percent.

Emmanuel Cortes developed a strategy to identify plant disease in his paper, an approach to detect plant disease using Generative Adversarial Networks. "Plant disease detection using CNN and GAN" [7]. For accurate extracting features and mapping output, background segmentations are required. Although utilizing GANs to classify diseases in plants appears to have promise, segmenting depending on the background had no effect on accuracy.

In their research "Convolutional Neural Network based Inception v3 Model for Wildlife Classification," Jyotsna Bankar recommended utilizing the inception v3 model to classify animals of various types [8]. Inception v3 has the capacity to both classify and categorize objects, which makes it useful in image classifiers. Prasanna Mohanty et al. devised a strategy for recognizing disease in plants by constructing a CNN model in their work "Deep learning based on Image- Based Plants Detection"[9]. The CNN model was taught to differentiate between healthy and ill plants from 14 different species. The test data provides 99.35 accuracy in the model. The model has a 31.4 percent accuracy when used on images obtained from reliable web sources; while this is superior to a simple model of randomly selected data, the most diverse collection of training data will help improve accuracy. Other types of method and otherwise neural network training also may improve accuracy, opening the way for anyone to detect plant disease. The numerous types of plant ailments and the various AI grouping algorithms that are used to recognize illnesses in various plant leaves.

3 Proposed system

3.1 Data sets

In order to make the machine to train deeply we are using some Deep learning Algorithms. So, we need large data sets to train the data. So we take the plant village dataset from kaggle.com. All of our

experiments utilize three separate Plant Village datasets. All of our experiments begin with the original color variant of the Plant Village set of data; the gray scale variant of the Plant Village set of data; lastly, we conduct all of our studies on a segmented Plant Village dataset, which removes all background information from the dataset because of the Plant Village dataset's regularized data collection process, thus removing any potential for inherent bias. A script customized for our database was used to automate segmentation. We are familiar with a widely used dataset for testing plant disease detection algorithms. We have overcome some plant disease plants like potato, tomato, pepper, rice and wheat. The data set we have taken had 15 directories and in each directory we had many files, nearly 4000 files. So, we have taken 80% data of files for training and 20% data of files to test the algorithm. So, for every type of crop we need data for most defected leaf, least defected leaf and medium defected leaf. So, we train and test data and calculate the accuracy of CNN model. Then we predict the actual output by using the images andget the name of the plant diseases.

3.2 Modules

This project's complete work is separated into five segments. Data collection, preprocessing, feature extraction, classification, and prediction are the steps involved. Data collection is the process of gathering, assessing, and interpreting correct findings for research using recognized validation procedures. Based upon the information required, many disciplines of research employ various data collection methods. The data was gathered from Kaggle's Plant Village dataset. The process of modifying or encrypting data so that it can be easily processed by a machine is known as data preparation. In other words, the algorithm can now swiftly deduce the features of the data. Using a technique known as feature extraction, it is possible to decrease the quantity of data needed to describe a large number of objects. A huge number of variables necessitates a lot of memory and processing resources, and it's difficult to generalize to a fresh sample. It's a catch-all word for methods of generating variable combinations to get past these issues while still accurately expressing the data. Convolutional Neural Networks are a type of Deep learning neural network. They're most commonly used to assess visual imagery and work in image classification behind the scenes. It's a strategy for classifying images. Passing an image to the programme predicts the output. The outcome of an algorithm after it has been learned on a historical dataset and applied to new data is referred to as prediction.

3.3 Architecture

The architecture describes how data will be accessed and interacted with in a variety of ways over the next few years, including by users and third parties.

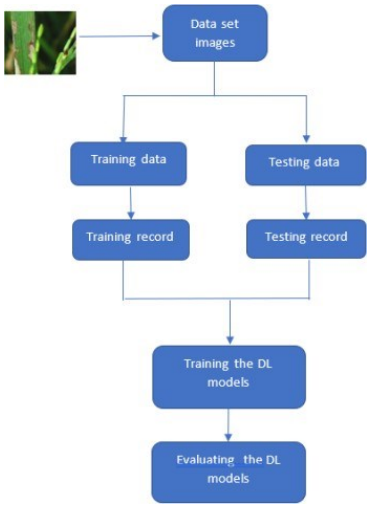

Fig. 2: System Architecture

3.4 Algorithm

Convolutional Neural Networks is a subset of deep learning, and have recently risen to prominence as the favored method [10]. CNN is the most extensively used image recognition classifier, with outstanding results in image processing and classification [11]. Deep learning was originally used to recognize plant diseases based upon leaf vein patterns [12]**.** To aid in the identification of a specific type of leaf illness, we suggest a methodology that leverages a Deep Neural Network, commonly known as a Convolution Neural Network. We'll use a Deep Neural Network to divide images and further classify diseases. We wanted the module to be designed in such a way that even someone with no programming experience could use and obtain information about plant disease. This framework can distinguish the disease of a large number of plant species, including Potato, Tomato, Rice, Bell Pepper, and Wheat. The photographs that are conveyed in the framework or present in the knowledge base complete the leaf's discovery. Second, characterization is carried out using images from the collection. A convolution core is defined in the convolution layer. This method shows how to retrieve a portion of the spatial information from a feature map. Before the data is transmitted to the full-connection layer, it is prepared using a combination of convolution operation and pooling layers. The neurons in this layer are totally linked with the neurons in the pyramidal tract's top layer.

3.5 CNN model steps

Conv2D: It is the layer that is used to split one image into numerous images. The activation function is what it sounds like.
MaxPooling2D: The largest value from the specified size matrix is pooled in this stage, and it is the same for another two layers
Flattern: After the image has been convolved, it is utilized to soften the image's dimensions.
Dense: It is the hidden state that is used to create thisfully linked model.
Dropout: The overfitting on the dataset is avoided here, and the output layer has only one neuron that determines which category each image belongs to called dense.
Image Data Generator: It is a process of resizing the image, shear the range, zooms and flip the image horizontally. It is an all possible orientation model.
Training Process: The function train_datagen_flow_from_directory is used so data from the train dataset directory should be prepared, The Target size specifies the image's goal size. To prepare test data for the model, use the test_datagen_flow_from _directory function. The model is created to fit the data into it, a fit generator is utilized. Other criteria used include steps per epochs, this measures the number of times the model will be run for the learned data.
Epochs: It indicates how many times the model has been taught in forward and backward passes.
Validation Process: Along with test data, validation data is supplied into the model.
In this Literature Survey we can conclude that. To detect the correct output accurately we can use the Convolution Neural Network (CNN) to get output more accurately.so we proposed this plant diseases identification using cnn.

4. Experiments and results

We used 3900 images from the dataset to test and train the data. We convert each image into an array, to resize the image the function convert_image_to_array is used. To get the default size we use DEFAULT_IMAGE_SIZE, then we build a model. 80% of the data is utilized to train the model, whereas 20% is used to test it. Tensor backend is used. We use batch normalization, maximal aggregation, and a reduction of 27% (0.25).Dropout is a control method that prevents the rectification of difficult collaboration data for training, reducing neural network readjustment. For summing neural network models, it's a very effective method. In our model, we use Adam's optimizer for optimization. Our network begins with the compiling the model and then the model.fit_generator function. Our goal is to add data, train-test data, and the number of epochs to train. We were using a value of 25 for epochs in this project then we plot the training and validation graphs to compare the performance and loss.

Fig. 3: Graph Showing Training and validation Loss

Fig. 4: Graph Showing Training and validation Accuracy

After that we evaluate the data, which gives the test accuracy as 98.7475 and we save the model. Now we test the model with predict_disease function, where we give the image path as its parameter. So we predicted output for blueberry_healthy, Potato_early_blight, tomato_target_spot, orange_Haunglongbing.

Fig.5 : blueberry_healthy Leaf

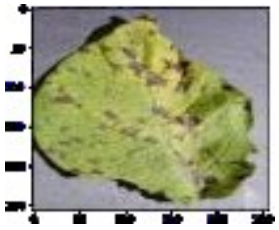

Fig.6 : Potato_early_blight Leaf

Fig.7 : Tomato_target_spot Leaf

Fig.8: Orange_Hauglongbin Leaf

Now we need to rebuild the model for getting output without any training, so here we imported some libraries and given the image to predict_disease function and get the name of the disease of the plant, for example we have taken corn_(maize)_Northern_leaf_blight

Fig.9: Corn_(maize)_Northern_leaf_blight

4. Conclusion and future works

This paper discusses, a Deep Learning computation, such as a Convolutional Neural Network, is used with the goal of detecting infections in yields. The model is mostly tested on certain plant species for specific types of plant diseases. Tensor stream and Kera's structures were used to create the model, and the framework runs on Android. The basic framework results reveal that, when compared to other models, the Mobile Net model performs better and provides greater accuracy in identifying illnesses. The number of plant classes and their diseases will be increased as part of the project. The model will also be improved by extending the preparation and testing bounds. In the future we can make some improvements to this model by adding some parameters to increase the accuracy and make the model faster with high accuracy. So that we can make the model to identify the plant diseases more correctly even with simple infection.

References

[1] Leaf disease detection using image processing 'Sujatha R*, Y Sravan Kumar and Garine Uma Akhil School of Information Technology and Engineering, VIT University, Vellore.

[2] Mokhtar, U.; Ali, M.A.; Hassanien, A.E.; Hefny, H. Identifying two tomato leaf viruses using a support vector machine. In Information Systems Design and Intelligent Applications; Springer: Berlin/Heidelberg, Germany, 2015; pp. 771–782.

[3] Kaur, S.; Pandey, S.; Goel, S. Semi-automatic Leaf Disease Detection and Classification System for Soybean Culture. IET Image Process. 2018, 12, 1038–1048. [CrossRef].

[4] Babu, M.P.; Rao, B.S. Leaves Recognition Using Back Propagation Neural Network-Advice for Pest and Disease Control on crops. Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.3632&rep=rep1&type=pdf (accessed on 1 May 2021)

[5] S.Arivazhagan, R. Newlin Shebiah, S.Ananthi, S.Vishnu Varthini. 2013. Detection of unhealthy regions of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int: CIGR Journal.

[6] Huu Quan Cap, Katsumasa Suwa, Erika Fujita, Satoshi Kagiwada, Hiroyuki Uga, and Hitoshi Iyatomi. A deep learning approach for on-site plant leaf detection. 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA).

[7] Santhosh Kumar S, and B. K. Raghavendra. Diseases Detection of Various Plant Leaves Using Image Processing Techniques: A Review. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS).

[8] Jyotsna Bankar, and Nitin R Gavai. Convolutional Neural Network based Inception v3 Model for Animal Classification. International Journal of Advanced Research in Computer and communication Engineering, May 2018.

[9] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," Biosyst. Eng., vol. 172, pp. 84–91, Aug. 2018.

[10] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Comput. Electron. Agriculture., vol. 147, pp. 70–90, Apr. 2018.

[11] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," Comput. Electron. Agriculture., vol. 127, pp. 418–424, Sep. 2016.

[12] Boureau YL, Le Roux N, Bach F, Ponce J, Lecun Y. [IEEE 2011 IEEE inter-national conference on computer vision (ICCV)—Barcelona, Spain(2011.11.6–2011.11.13)] 2011 international conference on computer vision—ask the locals: multi-way local pooling for image recognition;2011. p. 2651–8.

[13] Zeiler MD, Fergus R. Stochastic pooling for regularization of deep convolutional neural networks. Eprint Arxiv. arXiv:1301.3557. 2013.

[14] Kotamraju, Siva Kumar, et al. "Implementation patterns of secured internet of things environment using advanced blockchain technologies." Materials Today: Proceedings (2021).

[15] Bharathi, C. R., et al. "A Node Authentication Model in Wireless Sensor Networks with Locked Cluster Generation." Design Methodologies and Tools for 5G Network Development and Application. IGI Global, 2021. 236- 250.