REVIEW ON PLANT LEAF DISEASE DETECTION BASED ON MACHINE LEARNING

Pooja Tambe, Shreya Jangam, Vaishnavi Sinkar, Seema Shah, Prof.Kishor Pathak

Computer Engineering, Nutan Maharashtra Institute of Engineering and Technology

Savitribai Phule PuneUniversity, India

ABSTRACT

The artificial intelligence is based on automatic plant leaf disease detection and classification for quick and easy detection ofdisease and then classifying it and performing required remedies to cure that disease.Our goals is towards increasing the productivity of crops in agriculture. In this paper, we show how machine learning can be used for plant disease recognitionin the context of image classification. We have used a publicly available Plant Village dataset which has 10 classes of diseases. Hence the problem that we have addressed is a multi class classification problem. We have compared five different models including VGG16, ResNet50, InceptionV3,InceptionResNet and InceptionV1 as the backbones for our work. We found that VGG16 which uses skipconnections using a cnn layer archives the best result on the test set.

Keywords:—*Convolutional Neural Network (CNN), Transfer learning, image classification, Plant Leaf Diseases Detection, Adam Optimizer.*

INTRODUCTION

It is very important to get an accurate diagnosis of plant diseases for global health and well being. Agriculture plays a very important role in the economic growth of our Country. Accuracy and speed are the two main factors that will decide the success of the automatic plant leaf disease detectionand classification model. It is difficult for the naked eye of a human being to catch all sorts of problems with plant diseases. Also doing this time and time again is also laborious and unproductive work. In order to achieve accurate plant disease detection, a plant pathologist should possess good observation skills so that one can identify characteristic symptoms. An automated system designed to help identify plant diseases by the plant's appearanceand visual symptoms could be of great help.We are using transfer learning algorithm.Transfer learningis a machine learning algorithm. Disease detection in plants plays an important role in agriculture as farmers have often to decide whether the crop they are harvesting is good enough. It is of utmost importance to take these seriously as it can lead to serious problems in plants due to which respective product quality, quantity or productivity is affected. Plant diseases cause a periodic outbreak of diseases leading to large-scale death which severely affects the economy. These problems need to be solved at the initial stage, to save the lives and money of people. Automatic classification of plant diseases is an important research topic as it is important in monitoring large fields of crops and at a very early stage if we can detect the symptoms of diseases

when they appear on plant leaves. This enables computer vision algorithms to provide image-based automatic inspection. Comparatively, manual identification is labor intensive, less accurate and can be done only in small areas at a time. By this method, the plant diseases can be identified at the initial stage itself and the pest and infection control tools can be used to solve pest problems while minimizing risks to people and the environment.

MOTIVATION

The economic requirements though in radically different ways to those currently used this has given rise to many new chances to service. So they shouldbe tested via non-destructive techniques Leaves are delicate part of plant, The evaluation of agricultural harvest Classification is dynamic. The most important visual property is leafs texture and color. Hence, classification of leaf disease is necessary in evaluating agricultural produce, increasing market value and meeting quality standards. Identifying and taking further dealings for further diffusion of the diseases it is also helpful. The process will betoo slow, If the identification and categorization is done through physical techniques, we need the experts help sometimes it will be error prone and who are less available. The labor's classify based oncolor, size etc. if these quality methods are recordedinto automatic system by using appropriate programdesign language then the effort will be error free and faster.There are two main characteristics of plant disease detection machine-learning methodsthat must be achieved, they are: speed and accuracy.There is need for developing technique such as automatic plant disease detection and classification using leaf image processing techniques. This will prove useful technique for farmers and will alert them at the right time before spreading of the disease over large area. Solution is composed offour main phases; in the first phase we create acolor transformation structure for the RGB leaf image and then, we apply color space transformation for the color transformation structure. Then image is segmented using the K- means clustering technique. In the second phase, unnecessary part (green area) within leaf area isremoved. In third phase we calculate the texturefeatures for the segmented infected object. Finally, in the fourth phase the extracted features are passed through a pre-trained neural network.

RELATED WORK

There are various approaches to Plant detection. However, Plant authors have used various techniques to evade detection and therefore this has led to a needfor new and better approaches that can provide more accurate results.

The implementation of proper techniques to identifyhealthy and diseased leaves helps in controlling croploss and increasing productivity. This section comprises different existing machine-learning techniques for the identification of plant diseases.

CNN architectures in the identification of 26 different plant diseases. Different CNN architectures to identify58 different plant diseases, achieving high levels of classification accuracy. In their approach, they also tested the CNN architecture with real-time images. Forexperimentation purposes, We used the PlantVillage dataset and data-augmentation techniques to increase the data size. We reported better accuracy than that of a traditional machine-learning-based approach.
Pretrained CNN Model were used in to detect 3 different Tomato diseases from healthy-leaf images with modified hyperparameters such as minibatch size, max epoch, and bias learning rate. Six different pre- trained network (AlexNet, VGG16, VGG19, GoogLeNet, ResNet101 and DenseNet201)are there. To identify 10 different diseases in plants, and they achieved the highest accuracy rate of 97.3% using GoogleNet. A pretrained VGG16 as the feature extractor and multiclass SVM were used in to classify different eggplant diseases. CNN architecture was usedin to identify leaf diseases in plants, and an accuracy rate of 93.82% was achieved.A VGG16, ResNet, and DenseNet model was used.But We used VGG16 Model to identify plant diseases from the plant village dataset. To increase the dataset size, they used a double generative adversarial network (DoubleGAN), which improved the performance results.Here we are using Backend as a Pretrained cnn model and cnn model is inGoogle colab.Our Frontend is a web development and it is made up of streamlit.In transfer Laerning we are using VGG16 Model.

LITERATURE SURVEY

For proposed work to be better one following literatureis analysized for existing systems working and critically evaluated on some evaluation method to find shortcomings from them.

1. **Paper Name**: Detection of plant disease using threshold k-mean cluster and Ann algorithm.

Author's Name : Trimi Neha Tete. SushmaKamlu.

The research paper [1] gives in detail view of how to perform various techniques of segmentation for identification of different plant disease. Thresholding and K- means cluster algorithms are done to detect differentdiseases in plant leaf.

2. **Paper Name**: An Approach for identificationof infections in vegetables using Image Processing Techniques.

Author's Name: Ibelsha.N. Hariprasad. N.

So, to go with basic model architecture werefer a research paper [2] which has to approach towards optimizing an ideal Algorithm with efficient methods of feature extraction and classification techniques in identifications of diseases in vegetables.

3. Paper Name: Plant Leaf Disease Detection and Classification Based on CNN with LVQAlgorithm.

Author's Name: Melike Sardogan. AdemTuncer. Yunus Ozen.

The research paper [3] gives insights of how to presents a Convolutional Neural Network (CNN) model and Learning Vector Quantization (LVQ) algorithm based methodfor tomato leaf disease detection and classification.

4. Paper Name: Plant Health Analyser. **Author's Name:** Giriraja C.V. C.M.Siddharth Ch. Saketa M. Sai Kiran. The research paper [4] provides the information how to implement an Automatic

plant health checking vehicle with the help ofMATLAB where a rover keeps moving in a nursery capturing the image of the leaves.

5. Paper Name: Leaf Diseases Detection and Classification using image processing. **Author's Name:** Prakash M. Mainkar, Shreekant Ghorpade, Mayur Adawadkar Agriculture is the main of the Indian economy. Almost 70% people depend on it &shares major part of the GDP. Diseases in crops mostly on the leaves affects on the reduction of both quality and quantity of agricultural Products.

EXISTING WORK

A lot of research has been done in the last decadeon plant disease detection using deep learning and computer vision. Machine Learning approaches include traditional computer vision algorithms like haar, hog sift, surf, image segmentation, Support Vector Machines (SVM), using K-Nearest Neighbours (KNN), K-means and Artificial Neural Networks (ANN) have been successfully applied to a lot of different datasets. Deep Learning based plant disease classification models includes the use of a variety of CNN models such as AlexNet, GoogleNet, VGGNet etc. It is seen oftentimes asthe dataset size is not enough, multi class classification with a lot of classes requires careful hyperparameter tuning to avoid overfitting as the model could easily get stuck in a local minimum. This method uses transfer learning feature extraction techniques to identify the diseases in which the captured images are processed forenhancement first.The infected part of the leaf has been used for the classification purpose using a deep neural network using a softmax layer.To determine the defect and severity areas of plant leaves, feature extraction and classification. Finally classification is achieved using SVM uses a pre- trained convolutional neural network using 1.8 million images and uses a fine-tuning strategy to transfer learned recognition capabilities from general domains to the specific challenge of Plant Identification task. This work uses various transfer learning architectures and shows how hyperparameter tuning can be done to achieve the best results. Our work builds on top of this and we demonstrate state of the art results for this particularproblem.

DATASET

A public dataset is provided which contains 54,305 images of diseased and healthy plant leaves

collected undercontrolled conditions. The images cover 14 species of crops, including: apple, blueberry, cherry,grape, orange, peach, pepper, potato, raspberry, soy, squash, strawberry and tomato. It contains images of 17 basic diseases, 4 bacterial diseases, 2 diseases caused by mold, 2 viral diseases and 1 disease caused by a mite. Each class label is a crop-disease pair, and we make an attempt to predict the crop-disease pair given just the image of the plantleaf.
But we are using tomato leaf as a dataset.There are 10classes in leaf diseases detection.
Table 1 shows all the classes present in the PlantVillage dataset.

Table 1. Some samples of tomato leaf diseases

Class	Sample Image of tomato leaves
Healthy Leaf	
Leaf Mold	
Late Blight	
Mosaic Virus	

For training and testing purposes, we used the standard open-access PlantVillage dataset, which consists of 54,305 numbers of healthy- and infected-plant leaves. Detailed database information, the number of classes and images in each class, their common and scientific names, and the disease-causing viruses are shown in Table 1. The database contains 10 different classes of plant species with healthy- and disease-affected-leaf images. All images were captured in laboratory conditions. Table 1 shows some sample leaf images from the PlantVillage datasets . In our experiment, we used Tomato leaf formats of PlantVillage datasets.
First, we collect leaf images, and then with Preprocessing leaf images of the same datase. All leaf images were divided into two sets, a training set and the testing set. To evaluate performance, we split leaf images into three different sets, namely 80–20 (80% training images and 20% testing images), 70–30 (70%training images and 30%testing images), and 60–40 (60% training images and 40% testing images).

Table 2. Detailed description of PlantVillage Dataset with relative information.

Class	**Plant Name**	**Disease Name**	**Causes Virus Name**	**Type of Disease**	**No. of Images**
C1	Tomato	Healthy	-	-	1591
C2	Tomato	Bacterial spot	Xanthomonas perforans	Bacterial	2127
C3	Tomato	Early blight	Alternaria˙sp	Fungal	1000
C4	Tomato	Late blight	Phytophthora infestans	Fungal	1909
C5	Tomato	Leaf Mold	Lycopersicon	Fungal	952
C6	Tomato	Septoria leaf spot	Septoria lycopersici	Fungal	1771
C7	Tomato	Spider mites	Tetranychus spp.	Pest	1676
C8	Tomato	Target Spot	Corynespora cassiicola	Fungal	1404
C9	Tomato	Tomato mosaic virus	Tomato mosaic	Viral	373
C10	Tomato	Tomato Yellow Leaf	Begomovirus	Viral	5357

PROPOSED METHOD

We have used the concept of transfer learning forthe classification. The main advantage in using transfer learning is that instead of starting the learning process from scratch, the model starts from patterns that have been learned when solving a different problem which is similar in nature to the one being solved. This way the model leveragesprevious learnings and avoids starting from scratch. In image classification, transfer learning is usually expressed through the use of pre-trained models. A pre-trained model is a model that was trained on a large benchmark dataset to solve a similar problem to the one that we want to solve. We used cnn pre- trained models weights for our work.

A. Image augmentation techniques

The images are resized to 256 x 256 pixels, and we perform both the model optimization and predictions on these downscaled images. We used data augmentation like shearing, zooming, flipping and brightness change to increase the dataset size toalmost double the original dataset size. Data augmentation techniques are often used together with traditional machine learning algorithms or deep learning algorithms to improve the accuracy of classification. In this study, the image augmentation method was used by using the tenserflow machine learning library in Python. Width and height

change, cutting, zooming, horizontal turning, brightness, and filling operations were performed for normal class images. The image rotation degree was set to be randomly generated from 0 to 45.

In this study, the image augmentation techniques were applied only to normal images in order to balance the distribution of the samples over the classes. The number of normal samples in the dataset was increased from 1,583 to 4,266 by performing the image augmentation techniques. In this manner, the number of samples for each class was equalized. This equal distribution makes it possible to use all of the data instead of selecting random data during the training process. It is expected that this situation increases the accuracy ofthe training and positively affects the classification results. Image augmentation techniques are shownin Fig 1.

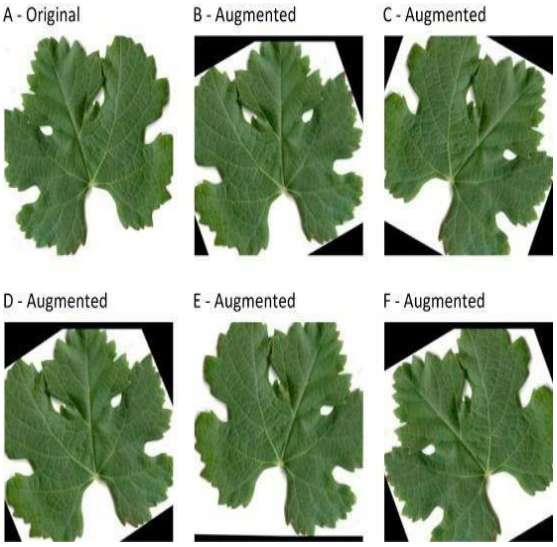

Fig 1:-Data augmentation

B. Optimization method

The main purpose of optimization methods is toupdate the weights at every stage until the best learning in CNN is realized. Each method performs an update process. In the Stochastic Gradient Descent (SGD) method, the weights update is performed in every iteration for each instance present in the training set. Because of this reason, it tries to achieve the goal as early as possible.

APPLICATIONS

Detection of plant disease through some automatic technique is beneficial as it reduces a large work of monitoring in big farms of crops, and at very early stage itself it detects the symptoms of diseases i.e. when they appear on plant leaves. Plant diseases and pests detection is a very important research content in the field of machine vision. It is a technology that usesmachine vision equipment to acquire images to judge whether there are diseases and pests in the collected plant images.

ADVANTAGES AND LIMITATIONS

A. Advantages of ML

Advantage of machine learning is to find the best solution, and we will find the best solution for you! advantages is a data-driven podcast that explains complex topics in a way that is easy to understand. Keep your devices safe and secure with the Advantagesplant detection software.Detection of plant disease through some automatic technique is beneficial as it reduces a large work of monitoring in big farms of crops.High processing speed and high classification accuracy. It is fast, simple and easy to program.It has no parameters to tune (except for the number of input).It is adaptable to any situation .It is powerful and it can adopt complex functions as well. advantagesis the best way to learn and teach. It's a learning systemwhere you can search for videos and lessons on anything, create your own content, and help others learn with you and can spot patterns and protect againstsimilar attack, could tell the difference between harmful and non-malicious files. The Advantage tool isa data flow dependencies visualization and editing application. It is a set of filters, modified data sets, anda visualization for the data flow dependencies of a model. The tool can help you to analyze and understand your model's data flow dependencies.

B. Limitations of ML

In order to obtain superior results in the detection of plant disease,the methods require a greater amount of data. Datasets are usually small and do not contain enough images, which is a necessity forhighquality decisions. This software prevents harmful software from running on your computer and is compatible with any antivirus installed on themachine. The machine learning disadvantage is a system that learns from its mistakes. It is an algorithm which employs clinical testing to different treatments, in order to find one that is more effective, for a given set of inputs. You can also use it to calculate some useful parameters for some other machine learning algorithms. To analyse data patterns, algorithms must be taught. When it comes to machine learning, there is a riskof using insufficient algorithms and creating limiting predictions.

FUTURE SCOPE

Other data sets can be added to improve accuracy, and more algorithms with greater performance can beadded to improve accuracy. One can further make a valid set, experiment with different methods. Development of a hybrid approach with two heads. Dynamic analysis will also be carried out, utilizing both automatic and boring methods, as well as a variety of dynamic instruments.Transfer Learning is asimple, but powerful probability-based classifier that automatically does some of the heavy-lifting for you. You can also use it to calculate some useful parameters for some other machine learning algorithms. Can be used by rural farmer to identify disease .Govt supported helpline number should be added to guide the farmer to manage the disease and pest control. Use of other algorithms

can be exploredto enhance the efficiency of the system in future.

CONCLUSION

We used PlantVillage dataset to train various CNN models. We have used transfer learning technique on various base models like VGG16 Model.VGG16has performed the best model.In this,we accurate artificial intelligence solution for detecting and classifying different plant leaf disease is presented which makes use of convolutional neural network for classification purpose. We were able to achieve more than 98.2% accuracy on validation and testset. Web application is built on the top of model, easy to use UI with simplified results are shown. Diseases in plants are a major threat to food supply worldwide. This paper demonstrates the technical feasibility of Machine learning using convolutional neural network approach to enable automatic disease diagnosis through image classification. Using a public dataset of 54,306 images of diseased and healthy plant leaves, a convolutional neural network is trained to classify crop species and disease status of 10 different classes containing, achieving an accuracy of 98.2% with residual network architecture. In this paper, a new approach of using machine learning methods was explored in order to automatically classify and detect plant diseases from leaf images. The developed modelwas able to distinguish between healthy leaves and different diseases, which can be visually diagnosed. The complete procedure was described, respectively, from collecting the images used for training and validation to image augmentation and finally the procedure of training the deep CNN and fine-tuning. We summarized the final results and came to the conclusion that VGG16 achieves the highest accuracy.

ACKNOWLEDGMENT

The authors would like to thank the publishers and researchers for making their resource available and also, we are thankful to our teachers for their guidance. We would like to express our gratitude to our guide Professor Kishor Pathak for his encouraging support and guidance in carrying out this work. We express our sincere thanks to Nutan Maharshtra Institute of Engineering and Technology ,Pune for permitting us to take our work this further.

REFERENCES

1. Plant Leaf Disease Detection and Classification Based on CNN with LVQAlgorithm" by Melike Sardogan, Adem Tuncer, Yunus Ozen in 3rd InternationalConference on Computer Science and Engineering, 2018.
2. Sethy, P.K.; Barpanda, N.K.; Rath, A.K.; Behera, S.K. Deep feature based rice leaf disease identification using support vector machine. Comput. Electron. Agric.2020, 175, 105527.
3. Chen, J.; Chen, J.; Zhang, D.; Sun, Y.; Nanehkaran, Y.A. Using deep transfer learning for image-based plant disease identification. Comput. Electron. Agric.2020, 173, 105393.
4. Trimi Neha Tete, Sushma Kamlu, "Detection of plant disease using threshold k-mean cluster and ann algorithm", Convergence in Technology(I2CT) 2017 2nd International Conference for, pp.

523-526, 2017**.**

5. N. Belsha, N. Hariprasad, "An approach for identification of infections in vegetables using image processing techniques", Innovations in Information Embedded and Communication Systems (ICIIECS) 2017 International Conferenceon, pp. 1-6, 2017.
6. C.V. Giriraja, C. M. Siddharth, Ch. Saketa, M. Sai Kiran, "Plant health analyser", Advances in Computing Communications and Informatics (ICACCI) 2017 International Conferenceon, pp. 1821-1825, 2017.
7. Singh, J.; Kaur, H. Plant disease detection based on region-based segmentation and KNN classifier. In Proceedings of the International Conference on ISMAC in Computational **Vision and Bio-**Engineering 2018, Palladam, India, 16–17 May 2018; pp. 1667–1675.
8. Tan, J.W.; Chang, S.W.; Kareem, S.B.A.; Yap, H.J.; Yong, K.T. Deep Learning for Plant Species Classification using Leaf Vein Morphometric. IEEE/ACM Trans. Comput. Biol. Bioinform. 2018, 17, 82– 90
9. Kamavisdar, P., Saluja, S., & Agrawal, S. (2013). A survey on image classification approaches and techniques. International Journal of Advanced Research in Computer and Communication Engineering, 2(1), 1005-1009.
10. P.Revathi, M.Hema Latha, Classification Of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques , ISBN, 2012, 169-173, IEEE.
11. Zelin[2008] A Study on the Method of Image PreProcessing for Recognition of Crop Diseases, International Conference on Advanced Computer Control, 2008
12. Ms. Kiran R. Gavhale, Prof. Ujwalla Gawande, and Mr. Kamal O. Hajari, "Unhealthy Region of Citrus Leaf Detection using Image Processing Techniques," IEEE International Conference on Convergence of Technology (I2CT)
13. P. Revathi and M. Hemalatha, "Identification of cotton diseases based on cross information gain_deep forward neural network classifier with PSO feature selection," *International Journal of Engineering and Technology*, vol. 5, no. 6, pp. 4637–4642, 2014.
14. Y. Jia, E. Shelhamer, J. Donahue et al., "Caffe: convolutional architecture for fast feature embedding," in *Proceedings of the ACM Conference on Multimedia (MM '14)*, pp. 675–678, ACM, Orlando, Fla, USA, November 2014.
15. A. K. Reyes, J. C. Caicedo, and J. E. Camargo, "Fine-tuning deep convolutional networks for plant recognition," in *Proceedings of the Working Notes of CLEF 2015 Conference*, 2015
16. M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid- level image representations using convolutional neural networks," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 1717–1724, IEEE, Columbus, Ohio, USA, June 2014.
17. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: delving deep into convolutional nets."
18. Ferentinos, K.P.: Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145, 311–318 (2018).

19. Cortes E. Plant Disease Classification Using Convolutional Networks and Generative Adversarial Networks. Stanford University Reports, Stanford(2017)
20. Fuentes, A., Yoon, S., Kim, S.C., Park,D.S.: A robust deep-learning-based detector for realtime tomato plant diseases and pest recognition.Sensors 17(9), 2022 (2017).
21. "Plant Leaf Disease Detection and Classification Using Image Processing Techniques" by Prakash M. Mainkar, Shreekant Ghorpade, Mayur Adawadkar in IJIERE Vol. 2, Issue 4, ISSN-2394-5494 , 2015
22. Aparajita, Rudransh Sharma, Anushikha Singh, Malay Kishore Dutta, Kamil Riha, Petr Kriz, "Image processing based automated identification of late blight disease fromleaf images of potato crops", Telecommunications and Signal Processing (TSP) 2017 40th International Conference on, pp. 758- 762, 2017.