IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

Prediction of Tomato Leaves Disease Using Ensemble Learning Algorithms

[1*] T.Ravi Kumar, [2] Kalyana Kiran Kumar, , [3] Suneelgoutham Karudumpa,[4] Ch.Rajasekhara Rao, [5*] Balamurali Pydi

[1] Dept of CSE, Aditya Institute of Technology and Management, Tekkali, AP, India.
[2] Dept of EEE, Aditya Institute of Technology and Management, Tekkali, AP, India.
[3] Dept of EEE, Aditya Institute of Technology and Management, Tekkali, AP, India.
[4] Dept of ECE, Aditya Institute of Technology and Management, Tekkali, AP, India.
[5] Dept of EEE, Aditya Institute of Technology and Management, Tekkali, AP, India.
[*]Corresponding Author: Balamurali Pydi balu_p4@yahoo.com

Abstract

Tomato is an important vegetable crop worldwide, and its production is threatened by various diseases. Major declines in crop yield and quality can be prevented by early diagnosis and detection of these diseases. Deep learning techniques have shown promising results in automatic detection of tomato leaf diseases. This survey paper presents a comprehensive overview of recent research on tomato leaf disease detection using deep learning. We summarize the datasets, architectures, and evaluation metrics used in the literature, and also find out some important challenges in upcoming days. But because of different leaf diseases as mosaic virus, bacterial spot, late blight, yellow leaf curl virus, etc., the quality and yield of tomato crops decline. Therefore, we are suggesting a deep learning-based system employing Resnet 152v2 and MobileNet v2 to detect the disease in tomato leaves, which makes use of many techniques to attain a decent crop production.

Keywords:-SupportVectorMachines;AlexNet;ResNet;LogisticRegression;K-NearestNeighbor(KNN); Convolutional NeuralNetwork(CNN); Naïve Bayes ;Decision Tree; Random Forest; Shuffle Net; Mobile Net.

1. Introduction

As a result of the large number of people who devote their lives to the agriculture sectors, India is known as the land of agriculture. The most widely grown crop on earth, the tomato maybe found in every kitchenin a variety of ways, regardless of the cuisine. India has 2^{nd}rankedin the production of tomato. Tomatoes biologically named is Solanum lycoperiscum grows mostly on well-drained soil is rich in vitamins like vitamin C, minerals and in their fields, nine out of 10 farmers cultivate tomatoes. To use fresh tomatoes from their garden and enjoy wonderful cuisine, many gardeners also favor growing tomatoes in their gardens. In general agriculture is becoming as backbone for Country's GDP, there is a need to identify the diseases which are affecting the yield of production. As farmers struggle a lot for proper crop production due to multiple diseases affecting the plant leaf so there is a need to detect the disease at early stages. Thus, detection of the plant leaf diseases in earlier stage also beneficial for Indian economy. So, to detect the diseases we have used deep learning-based approach. For disease classification and identification, a Convolution Neural Network-based method is employed. The experimental findings demonstrate the effectiveness of the suggested model compared to the trained model.

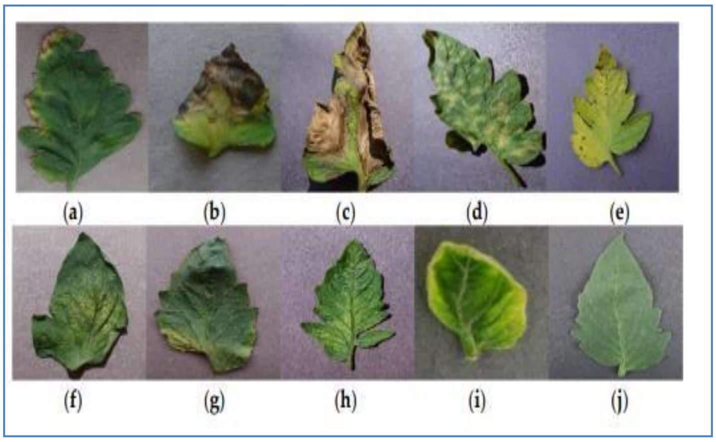
IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

2. Dataset

In[12], The 11,000 images of tomato leaves in the dataset for this study are divided into 10 groups. There are10 categories, 9 of which are diseased and 1 of which is healthy. Each category contains the same number of images which implies 11000 images in total.11,000 total images are used, 10,000 for training and 1,000 for validation. In each category, 1000 images are utilized for training and 100 images are used for validation. Each image in the collection are all 256x256 pixels in size, and there are no missing images. From the Plant Village dataset, images of tomato disease have been collected. Over 11,000 images of tomato leaves are present in the dataset. Here are a few instances of several tomato classes.



Sample dataset, where images organized by class.

The majority of tomato illnesses fall into nine categories:

(a) Bacterial_Spot (Xanthomonas vesicatoria): This is one kind of disease which can lead to death of that plant. This disease is very common in tomato plantations as this can be seen during the warm and wet climates.

(b) Early_Blight(fungus Alternaria solani): Itis a fungal pathogen .The pathogen causes "bullseye"-shaped leaf spots that are distinctive, as well as stem lesions and tomato fruit rot.

(c) Late_Blight(Phytophthora infestans): Any of the tomato plant's above-ground parts may contain it. Infected leaves typically have regions of dead tissue that range in colour from green to brown and are surrounded by a palegreen or greyborder. Late blight infections can seem water-soaked or dark brown in colour during extremely humid and damp conditions, and they are frequently described as looking greasy.

(d) Leaf_Mold (Cladosporium fulvum): It a non-obligate pathogen. Humid and chilly environments are likely to promote the pathogen's growth. This disease frequently manifests as symptoms on foliage, and it affects both the adaxial and abaxial surfaces of the leaves. The illness first affects the older leaves, then it advances to the younger leaves.

(e) Septoria_Leaf_Spot(fungus Septoria lycopersici): It is a fungal pathogen. Although this fungus can affect tomatoes at any stage of growth, symptoms typically first occur when plants are setting fruit on the older, lower leaves and stems. Thecalyx, petioles,stems, andleaves all occasionally show symptoms in addition to the usual leaf locations.

(f) Spider_Mites(floridana): It causes whitening or yellowing of leaves which dry out eventually, the

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

plant may die within a span of 3-5 weeks.

(g) Target_Spot(fungus Corynespora): It is a fungal pathogen. The disease is recognized by pits on the fruit and leaf damage that appears as target-shaped dots with light centers and dark borders. Tropical and subtropical regions of the world are affected by this illness, which affects tomatoes.
(h)Tomato_Mosaic_Virus(Tobamovirus): Thisvirusisplantpathogenic.Showsmottlingwithalternating yellowishanddarkergreenregions;thelatterisfrequentlythickerandelevated,givingtheappearanceofa blister. The seed may carry the virus and spread it. It just takes a few seedlings to become infected for the virus to begin spreading quickly. During everyday tasks, it can also spread to workers' hands, clothing, and infected instruments.

(i)Tomato_Yellow_Leaf_Curl_Virus(genusBegomovirus): It is a DNA virus that belongs to the family Gemini viridae and the genus Begomo virus. It is mostly destructive disease and found in tropical andsubtropical regions causing severe economic losses.

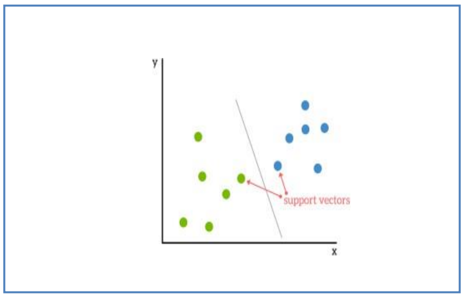
(j) Healthy_leaf: a leaf is not affected by any disease. It contains a mix of vitamins, minerals, proteins, antioxidants and other nutrients.

1. AlgorithmsandTechniquesUsed

The following are the algorithms or models which are used in the proposed application.

A. Support Vector Machine

SVM, also known as the Support Vector Machine, is a well-known Supervised Learning method that can be applied to both classification and regression issues. Yet, classification issues are mostly addressed by it in machine learning. The main goal of the SVM method is to choose the optimum line or decision boundary for categorising an n-dimensional space so that additional data points can be added to the correct category fast in the future. A hyperplane is the ideal boundary that might be chosen. SVM selects the extreme vectors and points that assist in building the hyperplane.



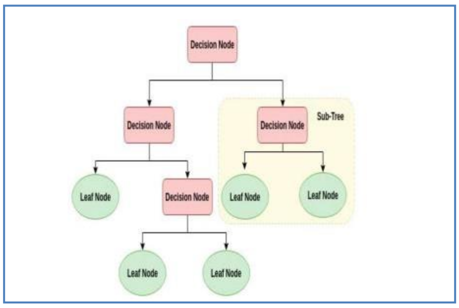
In[1],the data has 15,520 images divided into ten classes out of which nine are rotten and one is healthy. Diseases for each disease there is a dataset. The total images are divided into training which has 10,935 images and testing has 4585 images. The experimental results for identification of diseased tomato leaves by analyzing leaf images using machine learning and deep learning approach which can help pest identification and control in agriculture using SVM has achieved an training accuracy of 87.91% and achieved an accuracy of72.70% for validation accuracy. In[3], the dataset taken is “Plant Village” dataset which has 50,000images of various crops like tomato, grape, apple, corn and soya bean. From this, the data of tomato is used, that has ten classes including healthy leaf. The data divided into training and testing are in the ratio 4:1. The experimental analysis has achieved an accuracy of 91.0%and other metrics like precision, recall, F1 score were also evaluated. In [7], this

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

study introduced an approach that enables the classification of tomato disease in spreadsheets. The Plant Village image bank with open access was used. Three different tomato leaf diseases were identified from this data bank: late blight, mosaic virus, and tomato leaf spot Septoria. Each of the diseases was represented by 160 images. The experimental analysis shows that best results were obtained by SVM, which had a ratio of 80/20, a kernel rbf and 12 features. In [8], Two tomato leaf viruses are categorized and identified using the Support Vector Machine (SVM) technique with various kernel functions. TSWV and TYLCV infection datasets were used with 200 tomato splint images for both the training and testing phases. The experimental analysis shows that accuracy of 90% was achieved.

B. Decision Tree

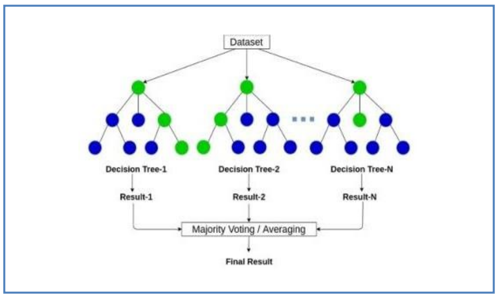
The most common application of a decision tree, which is a method of supervised learning, is classification problems, though it can also be used for regression problems. It is a classifier having a tree-like structure, with core nodes providing dataset properties, branches signifying decision-making steps, and leaf nodes expressing the result. The Decision tree's two nodes are the Leaf Node and the Decision Node. Decision nodes have a range of branches and are used to make decisions, as opposed to Leaf nodes, which indicate the results of decisions and have no extra branches.



In[2],Thereare4diseasesof tomato leaves like Bacterial Spots, Mosaic Virus, Septoria Spots, Yellow Curl and one healthy leaf with training data of 300 images and 200 images for testing. The experimental results using Decision Tree algorithm shows that 90% of bacterial spot diseases, 100% of healthy leaves, 82.5% mosaic leaves, 92.5% of Septoria leaves, 85% of yellow-curl leaves are correctly specified. This leads to overall accuracy using decision tree classifier is 90%.

C. Random Forest

Random Forest is a popular machine learning technique from the perspective of supervised learning. Regression and classification problems can both be resolved using machine learning. It is based on the idea of ensemble learning, which is a method that combines a lot of classifiers to solve a difficult



ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

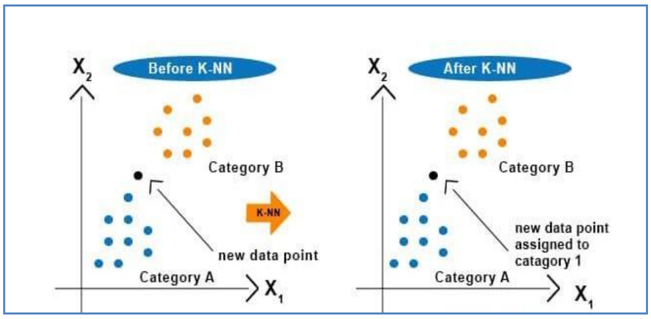
problem and enhance the performance of the model.

In[3], the dataset taken is "Plant Village" dataset which has 50,000 images of various crops like tomato, grape, apple, corn and soybean. From this, the data of tomato is used, that has ten classes including healthy leaf. The data divided into training and testing are in the ratio 4:1. The experimental analysis has achieved an accuracy of 82.7%and other metrics like precision, recall, F1 score were also evaluated. In [7], this study introduced an approach that enables the classification of tomato disease in spreadsheets. The Plant Village image bank with open access was used. Three different tomato leaf diseases were identified from this data bank: late blight, mosaic virus and tomato leaf spot Septoria. Each of the diseases was represented by 160 images. The results of the experimental analysis indicate that the accuracy is 90.7% when 100 trees and 18 features are used, respectively.

D. K-Nearest Neighbor(KNN)

K-Nearest Neighbors(KNN) is an efficient machine learning technique for classification and regression tasks. KNN belongs to the family of instance-based learning algorithms that analyses the instances of the training data to provide predictions for new instances. A distance metric (such as Euclidean distance) is used in KNN to identify the k closest neighbors to the new instance, and the predicted class or value of the new instance is the majority class or average value of its k nearest neighbors.

In [3], the dataset taken is "Plant Village" dataset which has 50,000 images of various crops like tomato, grape, apple, corn and soya bean. From this, the data of tomato is used, that has ten classes including healthy leaf. The data divided into training and testing are in the ratio 4:1.



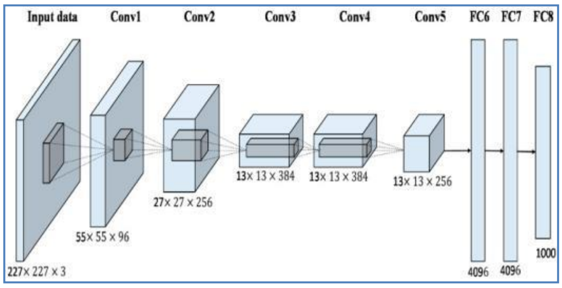
The experimental analysis has achieved an accuracy of 82.1% and other metrics like precision, recall, F1 score were also evaluated. In[7], this study introduced an approach that enables the classification of tomato disease in spreadsheets. The Plant Village image bank with open access was used. This data bank allowed for the discovery of three distinct tomato leaf diseases: late blight, mosaic virus, and tomato leaf spot Septoria. 160 photos were used to depict each ailment. With an accuracy of 85.3% for k = 3 and 12 features, it is discovered that K-NN performs better when the 80/20 relationship is taken into account.

E. AlexNet

Eight layers make up AlexNet: two fully linked layers, an output layer, and five convolutional layers.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

In order to avoid overfitting, the design makes use of the Rectified Linear Unit (ReLU) activation function and dropout regularization. Using graphics processing units (GPUs) for training allowed for faster processing of massive amounts of data, which was one of AlexNet's significant achievements. Moreover, the architecture introduced the idea of overlapping pooling layers, which assisted in lowering overfitting and raising accuracy.



In [3], the dataset taken is "Plant Village" dataset which has 50,000 images of various crops like tomato, grape, apple, corn and soya bean. From this, the data of tomato is used, that has ten classes including healthy leaf. The data divided into training and testing are in the ratio 4:1.The experimental analysis has achieved an accuracy of 92.7% and other metrics like precision, recall,F1 score were also evaluated. In [4] the dataset consists of450 images of blurred, occluded, scaled and distorted that include 9 classes of test and train images which are taken in the 3:2 ratio. Here KNN is used as a classifier and the experimental analysis has achieved an accuracy of 76.1%. In [10], the experimental analysis shows that the accuracy achieved is 81.20%.

F. Shuffle Net

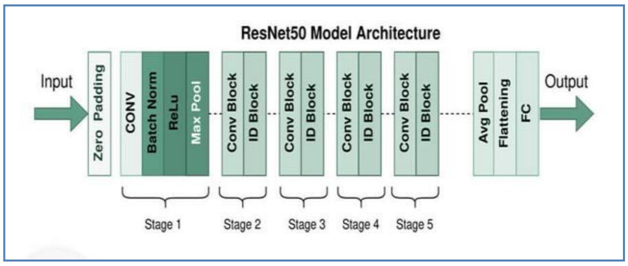
Shuffle Net uses a group convolutional operation to reduce the number of parameters and computation needed in the network. The group convolution operation divides the input feature map into a number of groups and applies a separate convolution to each group, then concatenates the outputs. Compared to a typical convolution process, this one requires less parameters and computational resources. Another key component of ShuffleNet is the channel shuffle operation, which is used to increase the network's representation capacity while keeping the number of parameters low. This operation shuffles the output channels of a group convolutional operation, allowing information from different channels to be combined in a more effective way. In[5],the data of diseased and healthy tomato leaves of 13112 images which are classified into 18 types are taken and the training and testing data is in the ratio of 4:1. The experimental analysis has achieved an accuracy of 83.68% for training data.

G. Resnet50

ResNet-50 consists of 50 layers, including convolutional layers, pooling layers, and fully connected layers. The architecture makes use of skip connections, which enable data to travel directly from one or more network layers to a subsequent one. As a result, the vanishing gradient issue is lessened and the network is able to learn more intricate characteristics. Additionally, ResNet-50 employs residual blocks, a kind of skip connection that adds a layer's input to its output rather than concatenating it. In place of learning the complete mappings, this enables the network to learn residual mappings, which

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

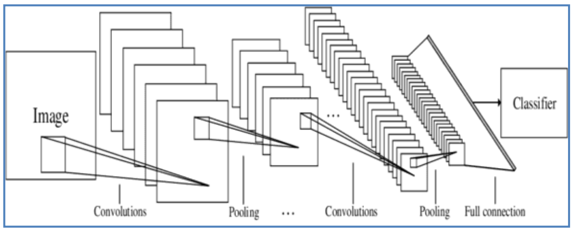
are the variations between a layer's input and output. This strategy has been demonstrated to increase training accuracy and speed.



In [5], statistics on healthy and sick tomato leaves of 13112 images which are classified into 18 types are taken and the training and testing data is in the ratio of 4:1. The experimental analysis has achieved an accuracy of 86.56% for training data. In [10], the experimental analysis shows that accuracy achieved is 87.73%.

H. LeNet

LeNet short for LeNet5's seven-layer architecture is made up of three fully connected layers, two convolutional layers, and two subsampling (pooling) layers. A 32x32 grayscale image is used as the input for LeNet-5. To extract low-level features, a collection of filters is convolved with the image first, and then subsampling is used to lower the dimensionality of thef eature maps. To extract more complex features, this procedure is repeated in the second convolutional and subsampling layer. Three fully linked layers that perform classification are then given the output of the second subsampling layer after it has been flattened. The output takes the ultimate form of a probability distribution over the classes of interest.



In[6],the dataset consists of 18160 images where 4800 used for testing and 13360 used for training. The experimental analysis has achieved an accuracy of 90% for 10 epochs.

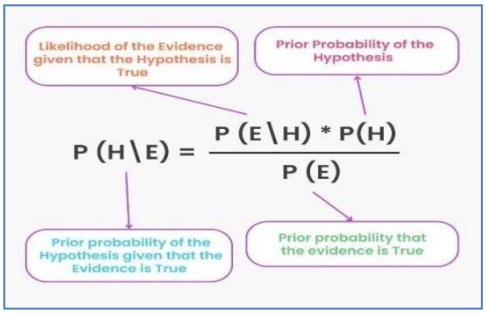
I. NaiveBayes

The Naive Bayes algorithm is a supervised learning algorithm that addresses classification issues by applying the Bayes theorem. The Nave Bayes Classifier is a straightforward and efficient classification technique that encourages the creation of quick machine learning models capable of making prompt predictions. Being a probabilistic classifier, it makes predictions based on the likely hood of an object. It is called naïve because it thinks that the incidence of one feature has nothing to do with the occurrence of other traits. For instance, if fruit is categorized according to its color, shape, and flavor, a red, spherical fruit with a mouthwatering flavor is known as an apple. As a result, each

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

trait, by itself, contributes to classifying it as an apple. It is called Bayes because it is based on the Bayes Theorem. When determining the likelihood of a hypothesis is given the relevant evidence, one uses the Bayes theorem, often known as the Bayes Rule or the Bayes law. It is driven by conditional probability.

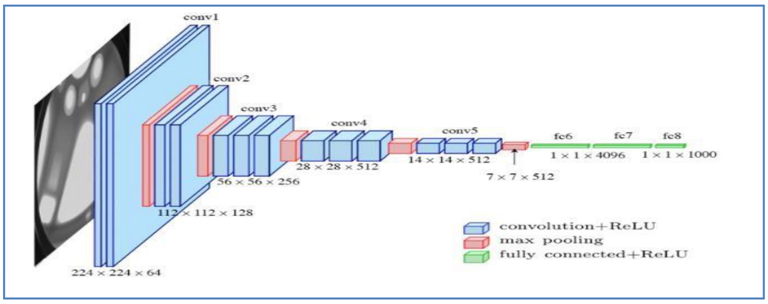


In [7], this study introduced an approach that enables the classification of tomato disease in spreadsheets. The public Plant Village image bank was consulted. This data bank allowed for the discovery of three distinct tomato leaf diseases: late blight, mosaic virus, and tomato leaf spot Septoria. 160 photos were used to depict each disease. The experimental analysis shows that with the 80/20 split and 18 features Naive Bayes performed better, producing an accuracy of 80.4%.

J. VGG16

It was one of the first CNN architectures to demonstrate excellent performance on image classification tasks and has since become a popular benchmark for image recognition.

VGG16 has a total of 16 layers, including 13 convolutional layers and 3 fully connected layers. The convolutional layers are arranged in blocks of two or three, with each block followed by a max-pooling layer. The filters in the convolutional layers have a very small receptive field of 3x3 pixels and are padded to preserve the spatial dimensions of the input. The VGG16 architecture has been pre-trained on the ImageNet dataset, which consists of over 1 million labeled images across 1,000 categories. The pre-trained weights can be fine- tuned on a smaller dataset for a specific image classification task. In [9], the deep VGG16 model trained with transfer learning, which achieves an overall accuracy of 90.4%



.

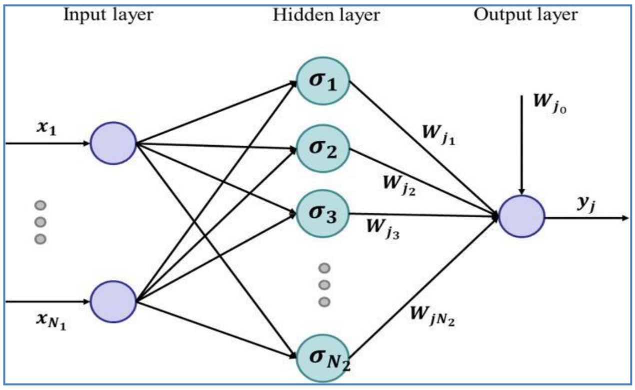
IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

K. InceptionV3

The Inception v3 architecture uses a novel module called "Inception module" that allows the network to learn multiple levels of abstract features. The module consists of a series of parallel convolutional layers of different sizes, followed by a pooling layer and a concatenation layer. This allows the network to extract information from different scales and resolutions of the input image. Inception v3 also incorporates other techniques to improve accuracy and reduce overfitting, such as batch normalization, regularization, and label smoothing[10].

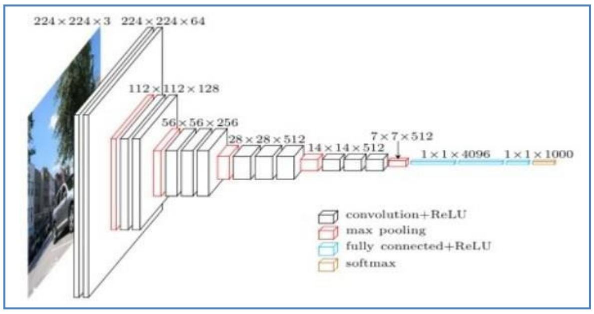
L. Extreme Learning Machine

A single hidden layer feed forward neural network's input weights and biases are initialized randomly in the ELM machine learning process. The Moore-Penrose pseudo inverse approach is then used to determine the output weights analytically, ensuring that they can be discovered rapidly and precisely. The network is ready to make predictions after the output weights have been established. ELM is superior to conventional neural networks in a number of ways, including quicker training times and better generalization performance. ELM has been effectively used in a number of applications, including speech recognition, time-series prediction, and image classification. In [11], the experimental analysis shows an accuracy of 71.8%



M. VGG11

Eight convolutional layers and three fully connected layers make up the total of 11 layers in the VGG11 algorithm. A 224x224 RGB image serves as the network's input, and a vector of class probabilities serves as its output. Small 3x3 filters with a stride of 1 and a padding of 1 are used in VGG11's convolutional layers, while 2x2 filters with a stride of 2 are used in the pooling layers. The VGG network family uses small convolutional filters with a modest stride and padding, which is one of its distinguishing characteristics. As a result, the network can acquire more precise visual information and make more precise predictions. But it also creates a very deep network, which can be difficult to train computationally. In[10], the experimental analysis shows that the accuracy of 86.7% is achieved.



Conclusion

Without the need for time-consuming feature engineering, deep neural networks have made it possible to implement promising solutions for plant pathology. Deep neural networks significantly enhance the accuracy of image classification. In order to address the issue at a nearly stage, precise and quick identification of plant leaf disease is required. For this, a computer vision method can be suggested to identify the leaf illness by taking pictures of the leaves and looking for signs of disease. In order to reach a conclusive decision that takes into account a range of leaf appearances, one can use a deep learning algorithm. Therefore, our suggested approach uses pretrained Deep Learning models to accurately identify illnesses in tomato leaves. Among various deep learning algorithms, we have used Resnet15v2 and MobileNetv2with softmax activation function. For Resnet15v2 we have achieved an accuracy of 96.79% which is our best model.

References

1. "Novel Computer Vision approach for identifying diseased Tomato leaves by Classifying Leaf Images for Pest Identification using ResNet152V2 Architecture." Karpagam, M. and Maheshwari, A. (2022).
2. Basavaiah, Audre, and Jagadeesh Anthony, Arlene. "Classification of tomato leaf diseases usingvarious feature extraction techniques." 633-651 in Wireless Personal Communications 115.1 (2020).
3. Tan, Lijuan, Jinzhu Lu, and Huanyu Jiang [3] are examples. "Classification of tomato leaf diseases using leaf images: A comparison of traditional machine learning and deep learning approaches." 542– 558 in AgriEngineering 3.3 (2021).
4. "Classification and identificationoftomatoleaf disease using deep neural network." Batool,Ayesha,et al. The ICEET 2020 conference is an international gathering on engineering and emergingtechnologies. IEEE, 2020.
5. Fenghua Huang, Hong, Huiqun, and Jinfa Lin. "Deep learning for tomato disease detection and classification." International Big Data and Artificial Intelligence Conference 2020
6. Tm, Prajwala,etal. "Tomatoleafdisease detection using convolutionalneuralnetworks."The eleventh international conference on modern computing (IC3) was held in 2018. IEEE, 2018.
7. J. C. Martinez-Perales, J. Cortes-Galicia, and B. Luna-Benoso. "Pattern recognition for tomato disease detection." 2020: 35–45 International Journal of Computing and Optimisation 7.1.
8. Usama Mokhtar, et al., "Identifying two of tomatoes leaf viruses using support vector machine." Information Systems Design and Intelligent Applications: Volume 1 of the Second International Conference on INDIA 2015 proceedings, published by Springer India, 2015.
9. Jianxin Wang, Wang, Guan, and Yu Sun. "Automatic deep learning-based disease severity assessment of plant diseases." 2017 edition of computational intelligence and neuroscience.
10. Wan, Hu, et al. "Plant disease classification using deep learning methods." The 4th

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 4, April 2022

International Conference on Machine Learning and Soft Computing Proceedings will be published in 2020.

11. Chuanqi Xie, et al. "Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging." (2015): 1–11 in Scientific Reports 5.1.
12. https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf