An International Conference-Innovation 2K23
in association with International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Plant Species Classification & Disease Detection Using Machine Learning

Krushnakant Bombe[1], Pranoti Chavan[1], Harshal Wagh[1], Prof. Ishwar Bharambe[2]
Prof . Minal Toley
[1]Student, [2]Assistant Professor
Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegoan, Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT

All living life depends on plants, and there are around 41 million different species of plants on earth which provide us oxygen, food and other essentials for survival of human beings. Excellent knowledge of plants is required to identify rare or some new plant species to improve the balance of ecosystem. The matching of plants with known taxa is called as plant identification that means classification of specific plants into known taxa by comparing definite characteristics. Standards and systems used to identify plants is a process that has evolved over centuries. Identification of plants is essential because it allows us find appropriate facts related to various species to serve certain types of applications.

Keyword:-Machine Learning, plant species identification, plant disease detection, image pre-processing, segmentation, extraction, CNN

I. INTRODUCTION

Plants play vital role to preserve natural resources. Plant species identification provides essential details about characteristics of plants and how plants are classified. It is very difficult to observe plant diseases manually. To do so lot of time and efficient knowledge of plant disease is required. Image processing and machine learning models are use to identify plant diseases.

Problem Definition

Identifying plant species and plant diseases correctly is very difficult. Lot of efforts and skills are required for identification, extensive knowledge of plants and research is required for early detection of diseases. Image processing is the best option used for plant species and diseases detection. image acquisition, image extraction, image segmentation, and image pre-processing methods are used for species and disease detection.

II. LITERATURE REVIEW

SR.NO.	PAPER/ AUTHOR/ YEAR	DESCRIPTION
1.	Machine Learning Based Plant Diseases Detection: A Review. Hilman F.Pardede, EndangSuryawati, DikdikKrisnandi	Research shows that CNN have replaced easy architectures such as SVM as industry leaders,Still there are many difficulties to overcome. The first important question arises is resistance to environmental conditions. And how to manage wide range of data and disease with small amount of data is secondary.
2.	Recognition of Plant Diseases using Convolutional Neural Network. G.Madhulatha,O.Ramadevi	To identify early stages of plant disease symptoms and classify disease based on symptoms deep learning technique is used. identification of diseases with the best accuracy of 96.56% proposed model makes use of deep CNN. This accuracy rate validates the performance of the model as

		an advisory
3.	A Deep CNN Approach for Plant Disease Detection. Fatma Marzougui, Mohamed Elleuch, MonjiKherallah	Deep learning approach gives very robust performance for variety of object detection problems. According to the results obtained, the system created offers better detection performance than those proposed in prior art
4.	AIBased Indigenous Medicinal Plant Identification. Anu Paulson, RavishankarS.	In this study, the accuracy of CNN, VGG16 and VGG19 database is compared. The proposed datasets are used for testing and training. CNN achieves an accuracy of 95.79%. VGG16 and VGG19 outperform standard CNNs with an accuracy level of 97. 8per and 97.9per respectively.

This system offers an automatic diagnosis of foliar diseases of banana trees. Convolutional neural networks (CNNs) have been successfully studied for plant disease detection and classification; however, due to the inherent inability of maximum clustering layers in CNNs, CNNs cannot capture the pose and orientation of objects. Given these shortcomings, they used a new model called Capsule Network (CapsNet). The test accuracy of the built model reached 95%, correctly detecting banana bacterial wilt, black leaf spot and healthy leaves. In terms of rotation invariance, it beats three types of CNN architectures: CNN models trained from scratch, LeNet5 and ResNet50.

This system offers an automatic diagnosis of foliar diseases of banana trees. Convolutional neural networks (CNNs) have been successfully studied for plant disease detection and classification; however, due to the inherent inability of maximum clustering layers in CNNs, CNNs cannot capture the pose and orientation of objects. Given these shortcomings, they used a new model called Capsule Network (CapsNet). The constructed model was tested with 95% accuracy, correctly detecting banana bacterial

III. SOFTWARE REQUIREMENT SPECIFICATION

The background-free image is then subjected to further image segmentation processing. To extract elements such as color, shape, and texture from a photo, multiple segmented images are used. For farmers who want to analyze plant growth quickly and inexpensively, image processing is a useful and essential tool for identifying and classifying plant diseases and species. Therefore, a system using image processing is proposed to identify diseases early and classify them using a CNN method. In the proposed system, leaf images with complex backgrounds are taken as input.

This project proposes a survey on the identification and classification of plant species and the diseases that affect them. The human eye is limited in its ability to identify certain plant species and leaf diseases. Therefore, the use of image processing and machine learning techniques can be used to detect plant leaf diseases and accurately identify their species. The dataset was used to obtain photos for this survey. In the pre-processing stage, the background of the image is removed by background removal technology.

Project Scope

Both inexperienced gardeners and trained experts can greatly benefit from an automated system designed to use the appearance and visual symptoms of a plant's leaves to help identify diseases and plant species. Precision plant protection technology has potential for development and improvement, and advancements in computer vision are likely to drive the market for precision agriculture applications.

User classes and Characteristics

All Users: With its powerful and user-friendly design, Plant Species and Disease Detection Software is suitable for all users with the most basic computer skills.

Assumptions and Dependencies

At some point during execution, the project must have access to the response data. Until then, the test data will be used to provide a demonstration for the demo. It is assumed that users are familiar with using a keyboard and mouse and an Internet browser. An internet browser is required as the application is a desktop application. Consumers must have a reliable Internet connection.

IV. OUR SYSTEM

Proposed Work Plants have a crucial role in human lives for many reasons. There exist more than hundreds of kinds of plants in the natural ecosystem and it can be very difficult to distinguish between them. A legitimate comprehension of plants is required in recognizing new or uncommon plant species. Such recognizable proof will, thusly, improve the medication business; balance the biological system just as the agrarian profitability and manageability by expanding the production level. The plant identification in real world would be more preferable if done automatically using the mobile devices. Hence this is a mobile application, where we can recognize the plant on giving leaf image as input. The Convolutional Neural Networks are utilized to separate the highlights of pictures where the feature selection is finished by the machine itself. The development of the CNN architecture and the depth(the number of CNN layers)are critical focuses that ought to be stressed since they are answerable for the recognition capacity of the architecture of neural systems. The output will be the details like the biological name of the plant, common name, location of the plant, a nutrient requirement needed by the plant and the medicinal value of the plant(if any).

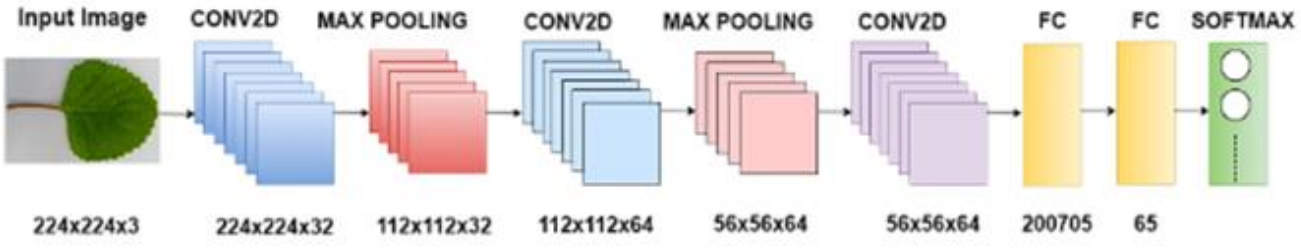

Fig. 4. Architecture of Convolutional Neural Network

These details can be stored in the database and we can assign an index value for each plant leaf. So this model will predict this index value upon fitting the image. This index value is searched in the database and the details

will be fetched and displayed. In this project, we are mainly focusing on Indian leaf with Medical values. Since our Android application works in realtime we have made our dataset by capturing the video of the plant and then converting it into frames. A 20s video is taken using a mobile camera and the video that was captured is broken into frames using online converters. We then processed the image to the required dimension so that it could fit into TensorFlow lite. At present we have collected 6 classes of medical plants with high medicinal value. For example, the plant Aloe Vera is found very useful in the treatment of hair loss in humans. The figure shown below is the images of some plants we have collected. Mobile-based automatic plant identification is essential to real-world socialbased ecological surveillance.INFOPLANT has a greater application in agriculture, education and the medical field. This project will help the farmers to raise the production level of crops by knowing the nutrient content of the plants. It will help in the proper usage of medicinal plants.INFOPLANT will help in creating awareness about the medicinal value of plants. Giving the leaf image as the input gives the medicinal value of plants if any present. Since it is being an android application, it will be user-friendly and free. It allows the students as well as professionals to identify an unknown plant without the help of a botanist. It will help in the economic growth of the country. It will predict this index value upon fitting the image. This index value is searched in the database and the details will be fetched and displayed. In this project, we are mainly focusing on Indian leaf with Medical values. Since our Android application works in realtime we have made our dataset by capturing the video of the plant and then converting it into frames. A 20s video is taken using a mobile camera and the video that was captured is broken into frames using online converters. We then processed the image to the required dimension so that it could fit into TensorFlow lite. At present we have collected 6 classes of medical plants with high medicinal value. For example, the plant Aloe Vera is found very useful in the treatment of hair loss in humans. The figure shown below is the images of some plants we have collected. Mobile-based automatic plant identification is essential to real-world social based ecological surveillance.INFOPLANTHas a greater application in agriculture, education and the medical field. This project will help the farmers to raise the production level of crops by knowing the nutrient content of the plants. It will help in the proper usage of medicinal plants.INFOPLANT will help in creating awareness about the medicinal value of plants. Giving the leaf image as the input gives the medicinal value of plants if any present. Since it is being an android application, it will be user-friendly and free. It allows the students as well as professionals to identify an unknown plant without the help of a botanist. It will help in the economic growth of the country.

V. CONVOLUTIONAL NEURAL NETWORK

1) Convolution Layer

The first block of CNN is that the convolutional layer. Primary purpose of convolution in CNN is to extract features from the input image. Convolution is a mathematical operation to combine two sets of data. The convolution is applied to the input data, utilizing a convolution filter to generate a feature map.

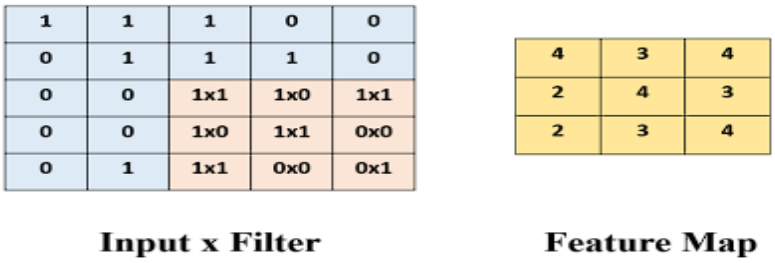

Fig. 5. Convolution Operation

In the figure given above, the input to the convolution layer is on the left side, for instance, the input image. On the right side is the 3x3 convolution filter, also called a kernel, often called so because of the filter shape.

2) Pooling Layer

Pooling performed to reduce dimensionality. It reduces the number of parameters, that minimizes the training time, and combats overfitting. Pooling layers downsample on each feature map independently while keeping important information. This layer also sums up the information obtained from the previous layer. Here max pooling is used, which takes the max value in the pooling window by sliding a window over its input. Pooling does not have any parameters. But it has two hyperparameters, that is, Filter(F) and Stride(S).

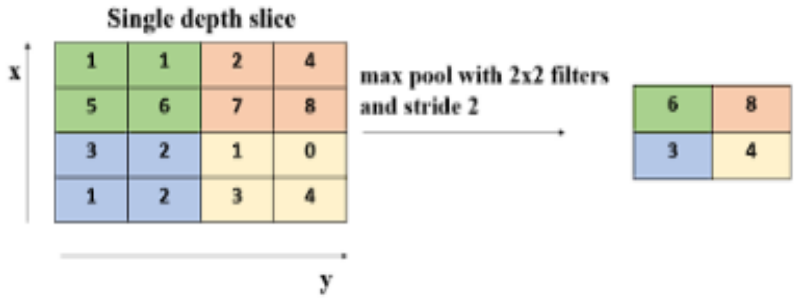

Fig. 6. Max Pooling

Fig. 6. Max Pooling Fig 6 shows the max pooling. Each window denoted by a different color. Pooling mostly is done employing 2x2 windows, stride 2, and no padding. And convolution completed using 3x3 windows, stride 1, with padding.

3) Fully Connected Layer

After the convolution and pooling layers, add fully connected layers. This layer contains weights, biases, and neurons. It will connect neurons in one layer to neurons in another layer. It can classify images between different categories by training. Since a fully connected layer expects a 1D vector of numbers, the output of the final pooling layer flattened to a vector, and that becomes the input to the fully connected layer. In this layer, the number of weights is n*m, with n inputs, and m outputs. This layer has (n+1)*m parameters with a bias for each output node.

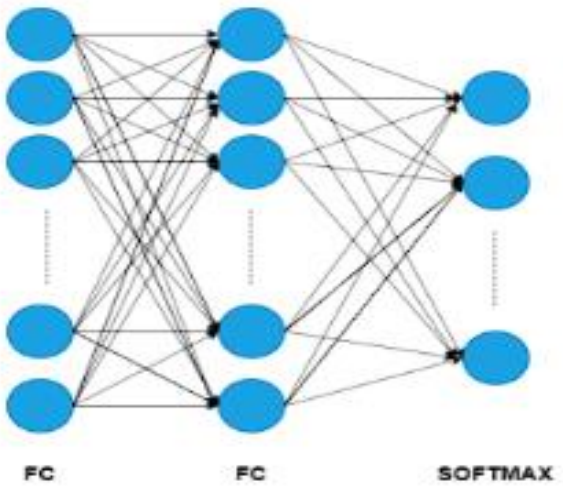

Fig. 7. Fully Connected Layer

VI. METHODOLOGY

The feature extraction task is done by Convolutional Neural Network otherwise called CNN. A convolutional neural framework involves input and an output layer, and many hidden layers. The hidden layers of a CNN routinely contain a progression of convolutional layers that convolve with an expansion or other dot product. The activation function is routinely a RELU layer and is subsequently trailed by additional convolutions, for instance, pooling layers and, fully connected layers and normalization layers, suggested as hidden layers because their sources of input and output are covered by the activation function and last convolution. The last convolu- tion, in this way, normally incorporates back propagation so increasingly the exact weight will be obtained as the outcome. Despite the way that the layers are conversationally alluded to as convolutions, this is simply by convention. Numerically, it is a sliding dot product or cross-correlation. This has hugeness for the indices in the grid, in that it impacts how weight is resolved at a specific index point. The input image is gone through various filters/kernels in the Convolution layer and each filter has a particular function. A few filters are for recognizing colors while others are for distinguishing the edges. Similarly, each filter has a particular task. At previous layers, essential features like lines, corners, and so forth will be found and at the last layers, a complete edge, and so on will be distinguished. At the last layers features such as stape of the item, and so forth will be found. So as the quantity of convolutional layers expands, more features will be extracted and afterward these features structures feature maps. After each convolutional layer, the pooling operation is done. Normal pooling tasks are average pooling and max pooling. We perform max pooling, which like a sliding window. A filter for instance, of 2*2 dimensions, will slide more than four pixels one after another and the maximum incentive out of those four pixels will be stored in one pixel of the following layer. In this manner, dimensions can be diminished and prominent features will be analyzed more. This procedure proceeds until a fully connected layer is reached. In a fully connected layer, every neuron is associated with each other neuron and afterward, the output is acquired from the last layer. In our model, the trained MobileNet model with our custom dataset. Below displayed images show the layers in our model as well as the input shape of those layers

VII. RESULT

In this project, we have used a web based Application that uses to identify the plant leaf and predict the details of the plant that the user has captured. We trained for 2 steps and for each step we have 12 epochs. Thus we have gained a prediction result with an accuracy of 99 percentage and validation accuracy of 95 percent.
The main advantage of our project is that details of the selected plant image could be predicted with a low-level Internet facility. Also, the user could easily predict the name of the plant without the internet. Shown below are screenshots of our Android Application.
As discussed earlier, first we will build a basic convolutional neural network from scratch, train it on the training image dataset, and evaluate the model. The architecture of CNN comprises of three convolutional layers with kernel size 3x3 and activation function as ReLU. This layer coupled with 2D max-pooling of size 2x2. The first 3 convolutional layers act as feature extraction. The output of the layer is input to the dense fully connected layer where the final prediction is done [8]. Finally, we will leverage the pre-trained models VGG16 and VGG19 which is already trained on a huge dataset with a diverse range of categories to extract features and classify images. The training loss values, training accuracy values, validation loss values, and validation accuracy values at successive epochs of CNN, VGG16, and VGG19 are shown in Fig 8, 9, and 10 respectively. The progress of accuracy and loss functions, during the training process and validation process, are shown in the graphs. The blue line shows the training loss. The orange line shows training accuracy. The green line shows validation loss, and the red line shows validation accuracy.

VIII. ADVANTAGES

As this is the only way to detect diseases at an early stage, automated plant disease detection systems have a definite advantage in monitoring large fields.
The enhanced species distribution map can automatically identify photos of plant samples, which is useful for ecosystem research and ecosystem resource protection.

IX. DISADVANTAGES

Machine learning methods require more data to more accurately detect diseases and plant species. This is a drawback, as currently available datasets are often limited and do not contain enough images, which are crucial for drawing high-quality conclusions.

X. CONCLUSION

Applications of machine learning and data mining in agriculture. This study uses CNNs to classify plant species and detect leaf diseases. Even so, the classifier was effective in predicting other diseases.

XI. REFERENCES

[1]. S. D. Khirade and A. B. Patil ”Plant Disease Detection Using Image Processing” (2015).

[2]. S. C. Madiwalar and M. V. Wyawahare, ”Plant disease identification: A comparative study” (2017).

[3]. P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka and E. Hernandez,”Plant Disease Detection Using Hyper spectral Imaging” (2017).

[4]. S. D.M., Akhilesh, S. A. Kumar, R. M.G. and P. C., ”Image based Plant Disease Detection in Pomegranate Plant for Bacterial Blight” (2019).

[5]. B. Wang, D. Brown, Y. Gao and J. L. Salle, ”Mobile plant leaf identification using smart phones”(2013).