Plant Leaf Recognition and Disease Detection Using GoogLeNet

[1]Ayushi Patani, [2]Isha Pardikar, [3]Paluck Doshi, [4]Shreya Rodge, [3]Shyamal Khachane

[1]Student, [2]Student, [3]Student, [4]Student, [5]Student,
[1]Department of Information Technology,
[1]Pune Institute of Computer Technology, Pune, India.

Abstract : Manually detecting disease of a plant can be a wearisome and difficult task. Furthermore, the level of accuracy is something that unintentionally may get compromised. To overcome this issue, an automated leaf disease detection system can be consolidated, via the means of image processing. In this paper, we propose a system using GoogLeNet that scans the leaf, identifies its type, and further recognizes the disease based on image recognition. GoogLeNet is advisable as it can run even on low computational devices. Leaf disease detection using GoogLeNet achieves a good level of accuracy. Furthermore, there can be some error factors which can be corrected using a discussion forum. It will also store information regarding the disease and the discussion taking place. In this forum, forward-looking farmers could take an initiative to spread agricultural information by being an active part of the community, and also help others to do so. Thus, the use of this leaf disease detection system will reduce both the time and effort required.

***IndexTerms* - Leaf Disease detection, Image acquisition, Pre-processing, Features extraction, Classification, GoogLeNet, CNN, Discussion Forum.**

I. INTRODUCTION

Agriculture is an important sector of the Indian economy as its contribution to the total GDP is about 17 percent and about 70 percent of rural households depend upon it. There are multiple factors upon which the growth of crops is determined. If any of the factors change, the crop can get affected. Various leaf diseases like a bacterial leaf spot, mosaic virus, early and late blight, etc. could be detrimental to plant growth. The disease first affects the leaf, and then spreads to the plant which makes it necessary for plant leaf disease detection. Therefore, it is important that the diseased crops are detected in the early stages to undertake countermeasures. As the cultivated areas are large, it is not possible to check each and every plant, hence, a camera-equipped system is must which can capture images of leaves, classify them, and identify the diseased leaf.

There are solutions proposed using various algorithms with accuracy as good as 99 percent, but there is no consideration about the 1 percent error. We propose a discussion forum along with the plant leaf disease detection system where farmers can clarify their doubts and take an initiative to help others by sharing information and being an active member on the forum. Agronomists can also guide other farmers regarding plant disease solutions.

This study utilized a dataset of 54,306 images (both healthy and diseased) i.e. the Plant village dataset that contains 38 classes of 14 crop species like apple, corn, grapes, potato, sugarcane, and tomato. A total of 26 various diseases were identified in the study. After the data acquisition, images were first preprocessed which included image annotation and image augmentation which is required for better feature extraction. Various algorithms like Random Forest, CNN, AlexNet, and GoogLeNet were studied and compared based on accuracy. The algorithm that had the best accuracy was considered.

This paper is organized as follows Section 2, includes the literature survey wherein various algorithms were studied and compared. Section 3 is about the methodology-image processing, studying the algorithm, and classification. Section 4 highlights the need for a discussion forum, Section 5, describes the future scope while Section 6, concludes the paper.

II. LITERATURE SURVEY

[1]. P. Sharma et al used a dataset with 19 classes of images that were preprocessed to resize pixels for speeding up computations, noise was removed using Gaussian Blur and converted to HSV to separate image intensity. Segmentation was performed using K Means clustering and the further pixel value of the background was changed to black to improve accuracy. Classifiers like Logistic regression, K Nearest Neighbors (KNN), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) were tested, and CNN has the highest accuracy of 98 percent was taken into account.

[3]. R. M. Prakash et al achieved 100 percent accuracy in detecting and recognizing the plant variety by segmenting the leaves using the K Means algorithm. The features were extracted using the Gray-Level Co-Occurrence Matrix (GLCM) and classified using Support Vector Machine (SVM). Minimum accuracy of 0.9 and a maximum of 1.0 was obtained on the used dataset.

[4] Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification proposed a model, using CaffeNet by google, to recognize 13 different types of plant diseases out of healthy leaves. It has the ability to distinguish plant leaves from its immediate surroundings. They proposed an approach of using deep learning methods to automatically classify and detect plant diseases from leaf images. The experimental results on the developed model achieved precision of 96.3% on average.

[5]Plant Leaf Recognition Using a Convolution Neural Network proposed a new method to classify leaves using the CNN model, and created two models by adjusting the network depth using GoogleNet.This paper focused more on the shape of the image ,they labelled leaves as oval,linear,oblong etc and achieved great accuracy even with 30% defective leaves.Leaf samples were used from Flavia dataset and achieved an accuracy greater than 90%.

[6] Gavhale, Ms & Gawande, Ujwalla explained and compared various plant disease detection techniques. These techniques include K-means clustering, SVM, BPNN and SGDM. The detection methodology consisted of RGB image acquisition, image component segmentation, computation of texture features, and configuring the neural network for recognition of disease.

Shima Ramesh, et.al.[7] put forward a method named Histogram of an Oriented Gradient to extract features of an image. Classifiers like logistic regression, support vector machine, k-nearest neighbor, CART, random forest, and naive Bayes were performed and compared. Random forest accounted for the highest accuracy of 70.14 percent.

[8] Mohanty SP, Hughes DP, and Salathé M in their paper proposed a method based on deep convolutional neural networks that detects leaf diseases in assistance with smartphone technology. RGB images of leaves are converted to grayscale and the unwanted background is blurred or removed. Further the image is compared with healthy and diseased leaf images and the disease is detected (if diseased).

Anuradha Badage,[9] proposed Canny's edge detection algorithm for monitoring the crop and identifying the disease. RGB values of the monitored images were extracted and compared with threshold images. Histogram analysis and edge detection techniques were used to identify particular plant diseases if the threshold found is greater or less than the given value.

[11] Jeon W, Rhee S. Plant Leaf Recognition Using a Convolution Neural Network proposed that machine learning is used to automatically classify leaf types. This uses deep learning models. Deep learning is itself a self-learning technique used on large amounts of data, and recent developments in hardware and big data help to work on the project.

[12] Plant Diseases Recognition Based on Image Processing Technology Guiling Sun, Xinglong Jia, and Tianyu GengNankai University, proposed a new image recognition system based on multiple linear regression. Particularly, there are a number of innovations in image segmentation and recognition systems.

III. Comparison

III.I. Random Forest

Random Forest is a classification algorithm. It consists of many decision trees where trees are not only trained on different sets of data but also use different features to make decisions. When building each tree, it uses feature randomness and bagging to try to create an uncorrelated forest of trees. Random Forest works poorly, as it gives an accuracy of only around 70.14% after using a linear kernel.

III.II. Convolutional Neural Network

CNN (Convolutional Neural Network) is the far most complex deep learning model that is used to classify the diseases. As it is very complex hence, it requires good computational power as well. It is the most common neural network that is applied to image classification problems.

After research we found that CNN performed quite well on the Plant village dataset and gave an accuracy of 97% which is much better than the accuracy of random forest.

The only problem with CNN is that it is hard to apply to high-resolution images. So there was a need of optimized GPUs and improved performance with cut down training times.

All traditional attempts lean heavily on hand-engineered features, image enhancement techniques, and a host of other complex and labor-intensive methodologies to detect plant diseases using computer vision as they. Also, traditional approaches to disease classification via machine learning typically focus on a small number of classes usually within a single crop. To overcome this flaw, AlexNet and GoogLeNet were introduced.

III.III. AlexNet

The architecture of alexnet has eight layers: five convolutional layers and three fully-connected layers. There are some of the features used that are new approaches to convolutional neural networks such as ReLU Nonlinearity, Multiple GPUs, and Overlapping Pooling (CNNs traditionally "pool" outputs of neighboring groups of neurons with no overlapping). AlexNet performed extremely well showing an accuracy of 98.01% on our database.

III.IV. GoogLeNet

It is a 22 layer deep neural network. ImageNet classifies images into 1000 object categories using the pretrained network. Computational efficiency was considered while designing the architecture. It can even run on low computational individual devices. The goal has been achieved by the top accuracy of 99.35% within the PlantVillage data set of 54,306 images containing 38 classes of 14 crop species and 26 diseases. Thus, without any feature engineering, the model correctly classifies crop and its disease.

IV. Methodology

IV.I. Phases Of CNN For Identification Of Disease In Plant

IV.I.I Pre-Processing Phase

The images collected from different sources have different sizes and qualities. To obtain consistency, these images are preprocessed for better feature extraction. The images are used as data by the classifier of the deep neural network. This process of expanding the data is known as image preprocessing which involves removing and normalizing the intensity of images, removing low-frequency background noise, image masking portion and reflection. It also involves cropping the images, highlighting the interest region by making the square on leaves. Then in the collecting phase, the images with a resolution of fewer than 500 pixels are considered invalid and the images with higher resolution are chosen for further processing. For feature extraction, the images should possess all the necessary information. The image preprocessing involves two steps: image annotation and image augmentation.

IV.I.I.I IMAGE AUGMENTATION

The overfitting problem of machine learning is overcome by regularizing the deep learning system. It involves selecting the hyperparameters and using a large number of images for training. Data augmentation is a method that is basically used when there are less images in the dataset. The techniques like geometrical transformations i.e. resizing, rotation, horizontal flipping, crop, and intensity transformations involving color, brightness enhancement, and contrast are used to increase the image dataset.

IV.I.I.II IMAGE ANNOTATION

Image annotation involves generating the description of a picture automatically, which is a major component in various retrieval applications and image search.

IV.I.II FEATURE EXTRACTION

A Feature Extractor is selected based on conditions like the type of layers, number of parameters, etc. A higher number of parameters increase the system complexity and directly influence the speed, and results of the system. Each network is therefore designed in such a way that the complexity is reduced but accuracy is increased.

IV.I.III CLASSIFICATION

In this phase, the classification model is used to determine the class of the disease of the given image. This model is generally trained using learning algorithms on the training dataset and tested against examples with a known disease.

IV.I.IV TESTING AND EVALUATING

In this phase, the dataset which is prepared by image annotation, image augmentation, feature extraction, and classification is tested, and the results are further evaluated.

IV.II. UNDERSTANDING GOOGLENET

IV.II.I DEEP LEARNING MODEL USING GOOGLENET

The model we propose in this paper is GoogLeNet.This model is a pre-trained network model that classifies an image into a thousand different classes that are trained on the basis of a large number of parameters. A GoogLeNet has 22 layers deep architecture. The initial four layers are simple convolutional layers with pooling followed by the first and fourth layers. Global Average Pooling is used over fully-connected layers in GoogLeNet , hence, reducing overfitting by reducing the number of parameters. It uses sparse network structures to improve the overall disadvantages of overfitting and over - occupying computing resources provided by AlexNet. It also uses the pyramid model to increase the width using the “Inception Module.” The “Inception Module” uses dense components to approximate the optimal local sparse structure. The GoogLeNet structure uses nine inception modules . These blocks of “inception modules” use equivalent 1×1, 3×3, and 5×5 convolutions so that it can capture a large number of features. The spatial dimensions are affected by max convolutions. Though this model is time-consuming, accuracy is very high. When the input image is provided to the trained model, it predicts the label associated with it and gives the probable percentage as the output.

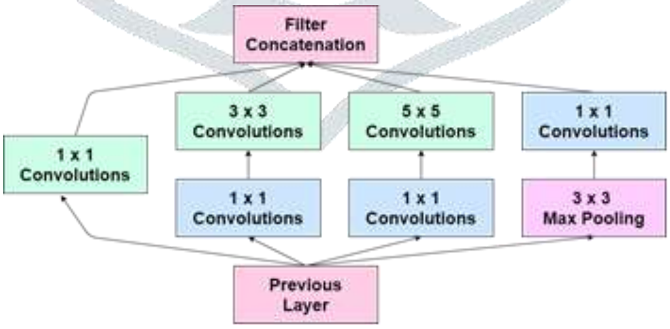

Fig 4.2.1.1: Inception Module with Dimensionality Reduction

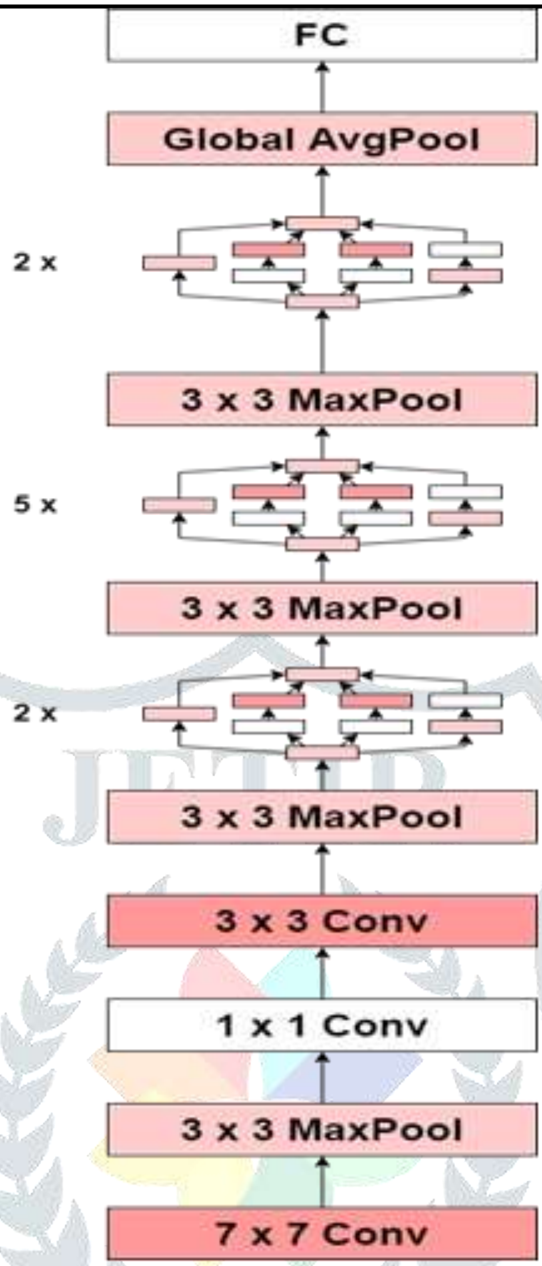

Fig 4.2.1.2: GoogLeNet Architecture

IV.II.II TRAINING GOOGLENET

The entire Plant Village dataset is divided into two parts -training set and test set. The division ratio usually used in any CNN model is 80/20. The method of Transfer learning is used to learn a new assignment. This method is simpler and faster than training a network from scratch. Transfer learning enables training even with fewer data. For re-training GoogLeNet to classify new images, the layers at the end of the network are replaced. Information about the combination of the extracted features into probability and class labels is included in the last layers. The last layers are fully connected and are of the same size as the number of classes. The problem of vanishing gradient is solved by storing these optimal values and adding the results of the auxiliary classifier using the backpropagation algorithm. This can result in stable learning results. At the end of the learning process, the auxiliary classifiers vanish and do not get utilized during the test stage.

Here, analysis of GoogLeNet's performance on the dataset is done by the transfer learning method of training a network. As inbuilt layers perform segmentation and feature extraction, it need not be done externally. Initially, the dataset images are loaded with the use of augmented image data stores. Next, from the pre-trained GoogLeNet network, a graph is extracted. Before training the network again, the final layers are replaced and initial layers are frozen. Finally, the network is trained using training data by augmentation and classification accuracy is calculated using validation data. As a result, we acquired an accuracy of 99.35%.

Fig 3.a. Apple leaf

Fig 3.b. Apple scab disease detected with an accuracy of 99.34%

Fig 4.a. Bell Pepper leaf

Fig 4.b. Bell Pepper bacterial spot leaf detected with an accuracy of 99.56%

Fig 5.a. Orange leaf

Fig 5.b. Orange huanglongbing disease detected with an accuracy of 99.7 %

Fig 6.a. Peach leaf

Fig 6.b. Peach healthy leaf detected with an accuracy of 98%

Fig 7.a. Potato leaf

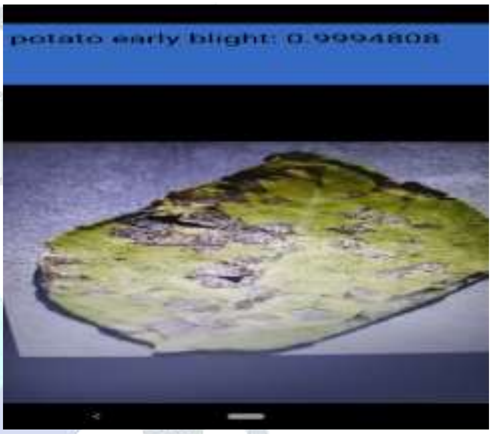

Fig 7.b. Potato early blight disease detected with an accuracy of 99.94%

Fig 8.a. Tomato leaf

Fig 8.b. Tomato leaf mold detected with an accuracy of 97.01%

V. Discussion Forum

Discussion forums are the preliminary form of social media. They are important from social perspectives as they mainly focus on providing a view on the topic being discussed or clarifying a doubt. They are often better for decision workflows.

Thus, the idea came up of introducing a virtual discussion forum in the application so the farmers can adapt to new technologies, identify more about plants and their related diseases, or learn more about agriculture. In this forum, forward-looking farmers could take an initiative to spread agricultural information by being an active part of the community and help others to do so. Agronomists can also guide other farmers regarding plant disease solutions. It is a process of consultation and dialogue between organizations of smallholder farmers and rural producers from all over the world. The Forum is rooted in and aims to strengthen effective partnerships and collaboration. The topics being discussed could be anything related to online sales of crops, aggregation, distribution, delivery, or peer discussion on Covid-19 challenges and opportunities. Messages are typically arranged by a thread—topic, date, and time—and allowed users to respond to a certain message or create a new message, or thread, of their own. It has a set of rules that discouraged malicious or inappropriate language and reserved the right to block any users that abused the forum. It will also have moderators, who viewed messages before they were posted to ensure they met the applications' standards.

Many farmers do not trust technology enough to make decisions about their crops based on their results. To ensure the results, the farmer can post their queries on the discussion forum. After further discussion, farmers can get the query resolved and progress towards repairing damage if needed.

VI. Future Scope

In recent years, farmers are facing a lot of economic challenges in terms of cost-effectiveness, productivity, and increasing labor shortage. Here the main monitoring aspect is not only to improve productivity but also to meet the demand of the growing population. Thus, on a large scale, monitoring of fields becomes quite a difficult task. So we propose a survey technique through the use of drones equipped with thermal and visible cameras. Drones can increase the productivity and quality of the farm. They also improve working conditions by the reduction of manual labor. They are available at affordable prices and are best suitable for monitoring large fields with corresponding geographic locations. They also give complete and clear pictures of fields. For example, multispectral and RGB cameras equipped drones are suitable for providing crop health conditions based on scanning the infrared portions of the crops. The main aim of introducing drones is to reduce the monotonous and time-consuming work of farmers. However, the measurement accuracy specifications are challenging to be addressed and several problems arise, the weather being one of them. For example, for drones, the wind influence, low GPS accuracy, in-flight stability, and image acquisition are major issues which are yet to be resolved.

VII. Conclusion

In this paper, we have proposed a method to classify plant leaves from the plant village dataset. We implemented a CNN model by adjusting the network depth using GoogleNet. It was observed that the model works perfectly fine even with low computational devices. We achieved an accuracy of 99.35% with the plant village dataset. Thus, without using any feature engineering, the model correctly classifies plants and its diseases. We also created a discussion forum where the farmers can post their queries. After further discussion, they can get their query resolved and progress towards repairing the damage, if needed. In the future, we may implement the proposed survey technique with the use of drones, equipped with thermal and visible cameras which can increase the productivity and quality of the farm.

References

[1] P. Sharma, P. Hans, and S. C. Gupta, "Classification Of Plant Leaf Diseases Using Machine Learning And Image Preprocessing Techniques," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 480-484, DOI: 10.1109/Confluence47617.2020.9057889.

[2] S. V. Militante, B. D. Gerardo, and N. V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IoT, Communication and Engineering (ECICE), Yunlin, Taiwan, 2019, pp. 579-582, DOI: 10.1109/ECICE47484.2019.8942686.

[3] R. M. Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-4, DOI: 10.1109/ICIIECS.2017.8275915.

[4] Sladojevic, Srdjan & Arsenovic, Marko & Anderla, Andras & Stefanović, Darko. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Computational Intelligence and Neuroscience. 2016. 1-11. 10.1155/2016/3289801.

[5] Jeon, Wang-Su & Rhee, Sang-Yong. (2017). Plant Leaf Recognition Using a Convolution Neural Network. The International Journal of Fuzzy Logic and Intelligent Systems. 17. 26-34. 10.5391/IJFIS.2017.17.1.26.

[6] Gavhale, Ms & Gawande, Ujwalla. (2014). An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. IOSR Journal of Computer Engineering. 16. 10-16. 10.9790/0661-16151016.

[7] Ramesh Maniyath, Shima & V, Vinod & M, Niveditha & R, Pooja & N, Prasad & N, Shashank & Ram, Hebbar. (2018). Plant Disease Detection Using Machine Learning. 41-45. 10.1109/ICDI3C.2018.00017.

[8] Mohanty SP, Hughes DP and Salathé M (2016) Using Deep Learning for Image-Based Plant Disease Detection. Front. Plant Sci. 7:1419. DOI: 10.3389/fpls.2016.01419.

[9] Anuradha Badage. (2018). Crop Disease Detection using Machine Learning: Indian Agriculture. Retrieved from International Research Journal of Engineering and Technology (IRJET).

[10] Sun G, Jia X, Geng T. Plant disease recognition based on image processing technology. J Electric Comput Eng. 2018;2018:1–7.

[11] Jeon W, Rhee S. Plant Leaf Recognition Using a Convolution Neural Network. IJFIS 2017;17:26-34. https://doi.org/10.5391/IJFIS.2017.17.1.26.

[12] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc.), 1097–1105. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., H.

[13] Caglayan, A., Guclu, O., & Can, A. B. (2013, September). "A plant recognition approach using shape and color features in leaf images." In International Conference on Image Analysis and Processing (pp. 161-170). Springer, Berlin, Heidelberg.

[14] H. Durmus, E. O. Gunes¸, and M. Kirci, Disease detection on the leaves of the tomato plants by using deep learning,

[15] Raghavendran.S. Plant Disease Detection System for Agricultural Application in Cloud Using CNN. International Journal of Computer Sciences and Engineering.