SEYBOLD Report
ISSN: 1533 - 9211
https://seyboldreport.net/

PLANT DISEASE DETECTION USING MACHINE LEARNING TECHNIQUES

Nihal

M.TECH Student (Department of Computer science & Engineering) Netaji Subhas University of Technology New Delhi, India,nihalsharma0606@gmail.com

Arvind Kumar

Professor (Department of Computer science & Engineering) Netaji Subhas University of Technology New Delhi, India er.arvindkumar@gmail.com

ABSTRACT-

In this study, we developed a classification model for plant disease detection using SOTA CNN architecture, specifically the VGG-16 architecture. We also used data and image augmentation techniques to improve the performance of the model. Compared with previous studies using the different SOTA CNN architecture for plant disease detection, we optimized the hyperparameters of the model to obtain a more accurate prediction. The results show that our model performs better prediction than previous studies and demonstrates the efficiency of transfer learning, data and image augmentation holders and hyperparameter optimization in improving the performance of deep learning models for plant disease diagnosis.

KEYWORDS-deep learning,transfer Learning,image augmentation ,classification

INTRODUCTION-

Plant diseases can cause significant economic losses in agriculture, and early detection and intervention can prevent the spread of diseases and minimize crop damage. However, traditional methods of plant disease detection rely on visual inspection by human experts, which can be time-consuming and may not always be accurate or reliable.

Machine learning and deep learning algorithms have shown great promise in automating the plant disease detection process. These algorithms can analyze large amounts of image data and identify patterns that are difficult for humans to discern, leading to faster and more accurate disease diagnosis.In particular, deep learning algorithms such as convolutional neural networks (CNNs) have been highly effective in image classification tasks, including plant disease detection. Transfer learning, which involves using a pre-trained deep learning model as a starting point and fine-tuning it for a specific task, has further improved the performance of these algorithms in plant disease detection.

Moreover, data and image augmentation techniques can improve the performance of deep learning models by increasing the size and diversity of the training dataset. By generating new images by applying transformations such as rotation, flipping, or cropping to the original images, data augmentation can help deep learning models generalize better to new, unseen images. Hyperparameter optimization is another critical aspect of deep learning model development. By systematically tuning the hyperparameters of a deep learning model, such as learning rate, batch size, and optimizer, researchers can achieve better performance and reduce overfitting.

Overall, the development of deep learning and machine learning algorithms for plant disease detection is a promising area of research with the potential to revolutionize agriculture by enabling faster and more accurate disease diagnosis and prevention. The study aims to develop a Deep Learning model with better test accuracy able to classify among the 38 plant classes using various deep learning techniques and hyperparameter optimization.

DATASET-

The dataset consist of 50,000 images of healthy and infected leaves images of 38 classes of plants.this dataset is an extension of the original plant village dataset which had around 55000 images of the same plants.this dataset curated with the help of crowdsourcing and subject experts.

Figure 1. Different type of healthy and infected leaves

LITERATURE SURVEY

A. Rastogi et al,[1] presents a novel approach to detecting and grading leaf diseases using computer vision technology and fuzzy logic. The proposed system consists of three main stages: pre-processing, feature extraction, and classification. In the pre-processing stage, the images of the leaves are enhanced using various image processing techniques such as contrast enhancement, noise removal, and histogram equalization. In the feature extraction stage, several features such as color, texture, and shape are extracted using various feature extraction techniques such as GLCM and Gabor filters. In the classification stage, fuzzy logic is used to classify the leaf images into healthy or diseased categories and to grade the severity of the disease and manage to achieve the prediction accuracy over 95 %.

H. Waghmare et al[2], presents a machine learning-based approach for the detection and classification of diseases in grape plants using opposite color local binary pattern (OCLBP) features. The proposed system consists of three main stages: image pre-processing, feature extraction, and classification. In the image pre-processing stage, the grape plant images are converted to gray-scale and normalized to reduce the effects of variations in lighting conditions. In the feature extraction stage, OCLBP features are extracted from the pre-processed images to capture the texture information of the plant leaves. Finally, in the classification stage, machine learning algorithms such as support vector machines (SVMs) and random forests (RFs) are used to classify the plant images into healthy or diseased categories and managed to achieve accuracy of around 89%.

P. B. Padol et al[3], proposes a grape leaf disease detection system based on a support vector machine (SVM) classifier. The proposed system consists of two main stages: feature extraction and classification. In the feature extraction stage, color and texture features are extracted from the grape leaf images using various techniques such as gray-level co-occurrence matrix (GLCM) and local binary pattern (LBP). In the classification stage, an SVM classifier is trained on the extracted features to classify the grape leaves into healthy or diseased categories.
The authors also discussed the limitations of their approach, including the requirement for manual feature selection and the limited ability to handle complex image datasets. The authors suggested that future work could focus on developing deep learning-based approaches that can automatically extract features and provide better performance on complex datasets.

K. kaur et al[4], Overall, the paper presents a SVM classifier-based approach for grape leaf disease detection using color and texture features. The proposed approach shows promising results and could be further improved by incorporating deep learning techniques for automatic feature extraction and better performance on complex datasets and managed to achieve prediction accuracy of 88.89%.

M.Chaudhary et al[5], proposes a fruit disease analysis system based on image processing techniques. The proposed system consists of four main stages: image acquisition, image pre-processing, feature extraction, and disease classification. In the image acquisition stage, the fruit images are captured using a digital camera. In the image pre-processing stage, the images are preprocessed to enhance the contrast and reduce the noise. In the feature extraction stage, various features such as color, texture, and shape are extracted from the fruit images. Finally, in the disease classification stage, machine learning algorithms such as k-nearest neighbor (KNN) and support vector machine (SVM) are used to classify the fruit images into healthy or diseased categories.Overall, the paper presents an image processing-based approach for the detection of plant diseases using machine learning algorithms such as SVM and decision tree. The proposed approach shows promising results and could be further improved by incorporating deep learning techniques for automatic feature extraction and better performance on complex datasets.

The paper proposes a method for analyzing and comparing infected leaves using image processing techniques such as Otsu thresholding and k-means clustering. The proposed method consists of four main stages: image acquisition, image pre-processing, feature extraction, and disease classification. In the image acquisition stage, the leaf images are captured using a digital camera. In the image pre-processing stage, the images are preprocessed to enhance the contrast and reduce the noise. In the feature extraction stage, various features such as color and texture are extracted from the leaf images. Finally, in the disease classification stage, Otsu thresholding and k-means clustering techniques are used to classify the leaf images into healthy or infected categories.

Overall, the paper presents an image processing-based approach for the analysis of infected leaves using techniques such as Otsu thresholding and k-means clustering. The proposed approach shows promising results and could be further improved by incorporating deep learning techniques for automatic feature extraction and better performance on complex datasets.

METHODOLOGY

1. Data Preprocessing: The first step in building any deep learning model is to prepare the dataset for training. In this project, we used a dataset containing 38 classes of plants With both healthy and unhealthy plant pictures to rain on. We divided the dataset into three sets: training, validation, and testing sets. We applied data augmentation to the training set images, specifically horizontal flip with a probability of 50%. This helps to increase the size of the training set and improve the model's generalization performance.

2. Transfer Learning: We used a variation of SOTA VGG16 architecture, a popular convolutional neural network (CNN) architecture, for our model. We replaced the original top layer of VGG16 with a batchNormalization with momentum of 0.99 and epsilon value of 0.001. Following this we added a dense layer with 256 units with L2 and L1 regularizer with l values of 0.016 and 0.006, and a Dropout layer with a rate of 0.45 to reduce overfitting, and then the second dense layer with 38 units (one for each class).

3. Model Compilation: We compiled the model using the Adamax optimizer, which is an extension of the Adam optimizer, with a learning rate of 0.001 and categorical cross-entropy as a loss function and accuracy as the metric.

4. Model Training: We trained the model on the training set with a batch size of 40 and a maximum of 40 epochs. To avoid overfitting and underfitting, we used early stopping and model checkpoint callbacks. The early stopping callback monitored the validation loss and stopped training if there was no improvement for 5 consecutive epochs. The model checkpoint callback saved the best model weights based on the validation accuracy.

5. Model Evaluation: We evaluated the trained model on the validation and testing sets to measure its accuracy and generalization performance.

6. Prediction: Finally, we saved the trained model and used it to predict the class labels of new images in the future. To predict the class label of an image, we first preprocessed the image using the same preprocessing steps as in the training set. We then passed the preprocessed image through the trained model and obtained the predicted class label.

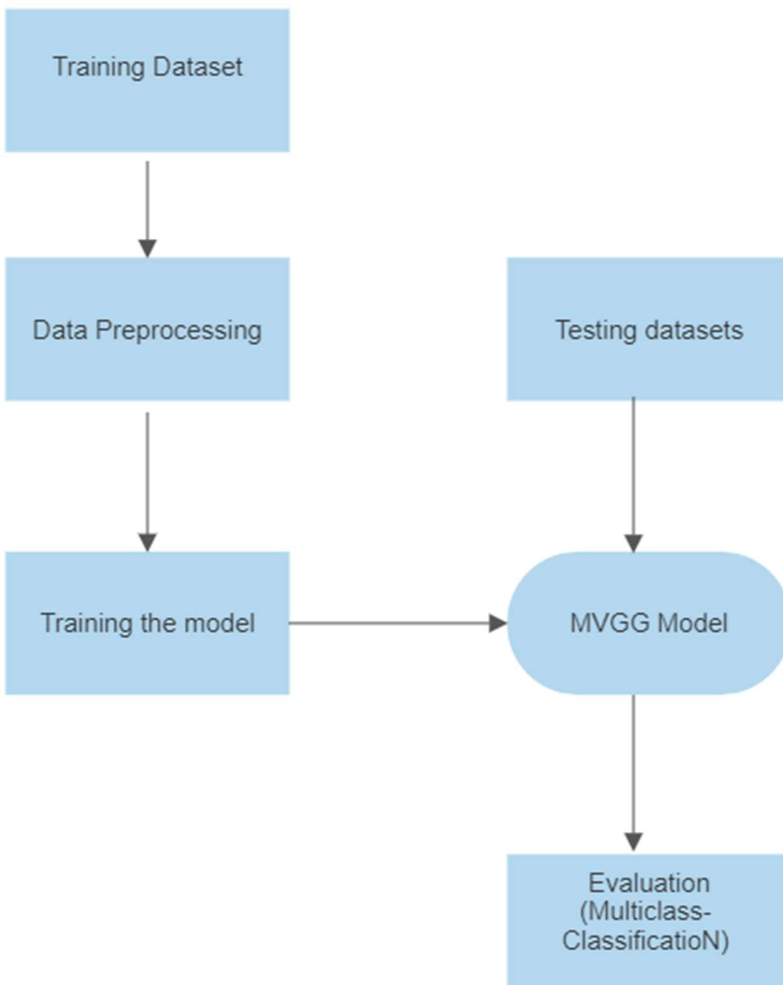

Figure 2 sequential flowchart of used methodology

SEYBOLD Report
ISSN: 1533 - 9211

WHY AdaMax optimizer-

By following the negative gradient of the function, the optimization approach known as gradient descent finds the minimum of an objective function. An issue with gradient descent is that there is just one step size (learning rate) used for all input variables. Extensions of gradient descent, such the Adaptive Movement Estimation (Adam) approach, use a separate step size for each input variable, but they may lead the step size to drop rapidly to incredibly low levels. AdaMax, an innovation in the Adam variant of gradient descent that builds on the concept of the infinite norm (max), may improve the effectiveness of particular optimizations.

CALLBACK FUNCTIONS-

Callbacks, in the simplest terms, are the unique utilities or activities that are carried out during training at specific points in the training process. Callbacks can assist you with a number of tasks, such as preventing overfitting, visualizing training results, troubleshooting your programme, generating logs, creating a TensorBoard, etc. TensorFlow offers a wide variety of callbacks, and you can utilize many. We'll examine the various callbacks available and look at some usage examples.

Overall, this methodology allowed us to build an accurate CNN model for classifying plant diseases using transfer learning with VGG16 architecture, data augmentation, and appropriate callbacks to avoid overfitting and underfitting.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 vgg16 (Functional)          (None, 512)               14714688

 batch_normalization (BatchN  (None, 512)              2048
 ormalization)

 dense (Dense)               (None, 256)               131328

 dropout (Dropout)           (None, 256)               0

 dense_1 (Dense)             (None, 38)                9766

=================================================================
Total params: 14,857,830
Trainable params: 14,856,806
Non-trainable params: 1,024
_________________________________________________________________
```

FIgure 3. Proposed variation of VGG16 architecture

CONCLUSION

Plant ailments reduce crop production, which jeopardizes the supply of food. As a result, when diseases are caught early, plants recover swiftly from them. This study demonstrates how convolutional neural networks and deep learning may be used to automatically diagnose diseases using picture categorization. A deep convolutional neural network was developed to categorize the 55,448 photos comprising healthy and diseased plant leaves from the publicly available PlantVillage dataset into 39 distinct groups, which contain 14 species and In order to automatically classify and detect plant diseases from leaf photos, a transfer learning algorithm was modified for this research to collect generic features from the vast plantvillage dataset.

The complete procedure was laid out, from collecting the photos used for training and validation until deep CNN training in the end. When used in conjunction with the AdaMax optimizer, VGG[16] achieved the highest accuracy rating of 99.79 on test data.

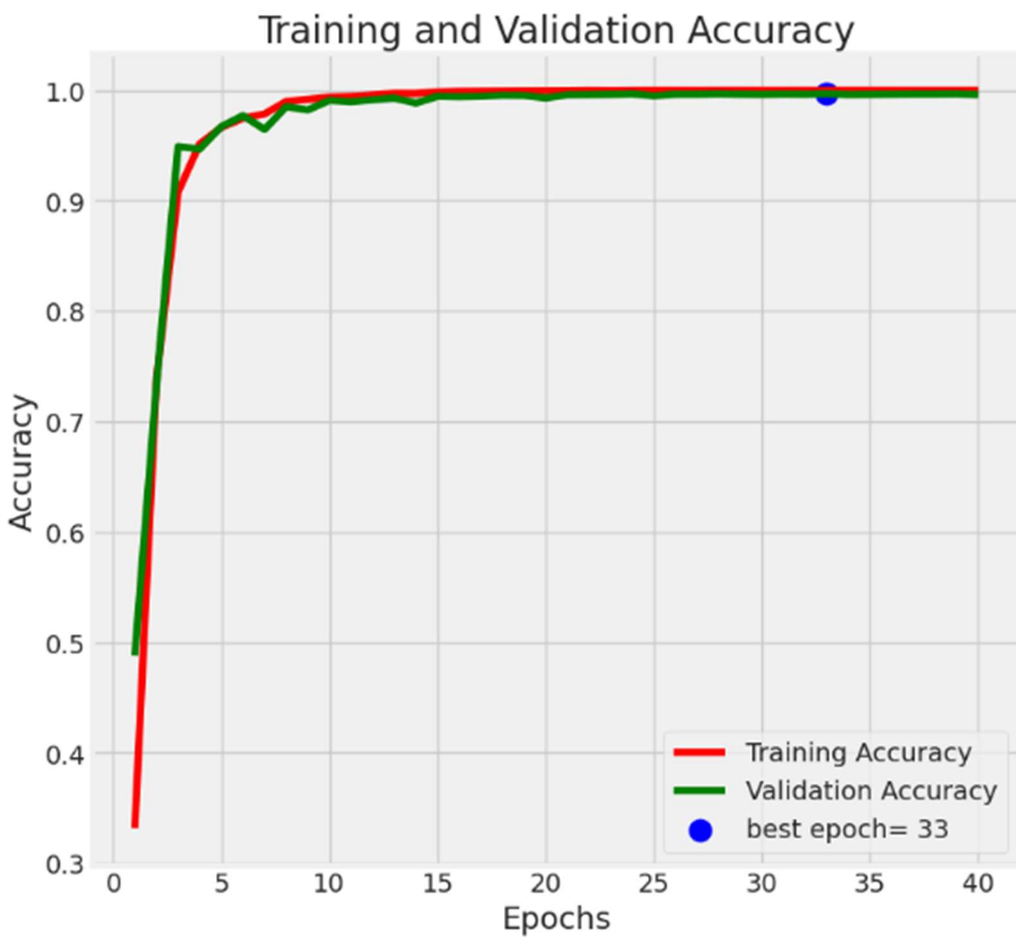

Figure 4 Training and validation accuracy loss w.r.t epochs

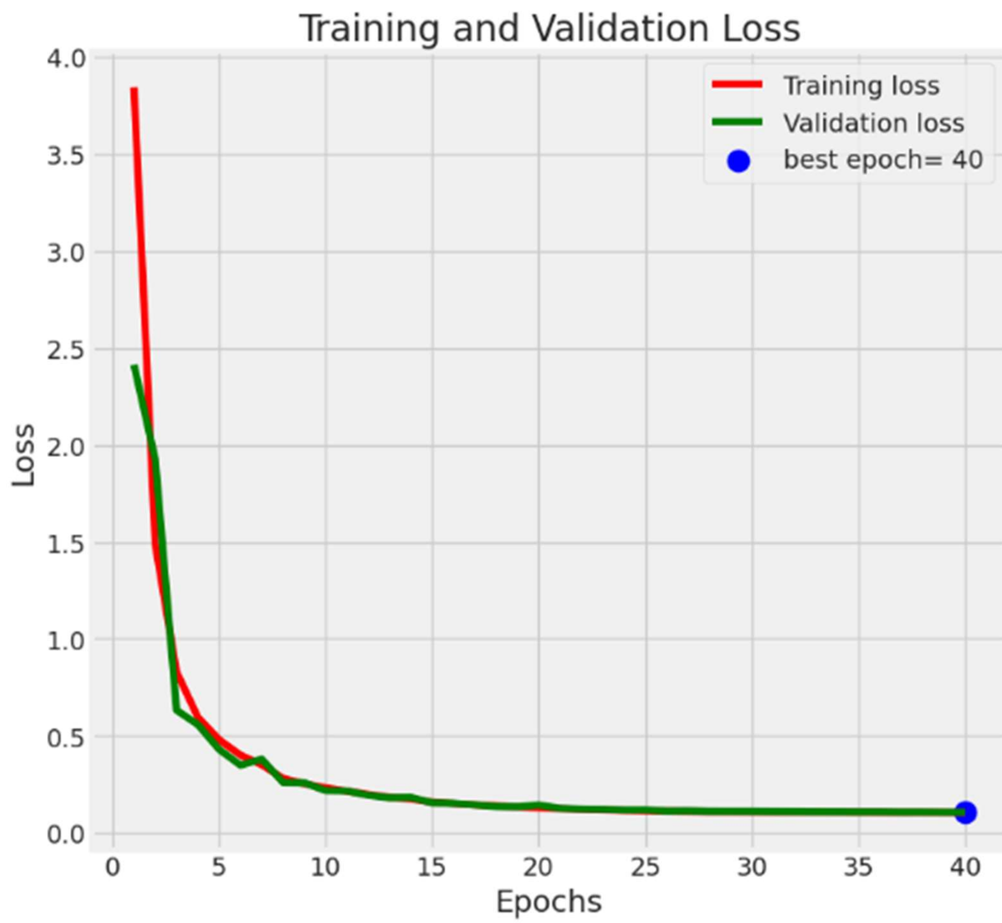

Figure 5 Training and validation loss w.r.t epochs

REFERENCES

[1] A. Rastogi, R. Arora, S. Sharma, "Leaf Disease Detection and Grading using Computer Vision Technology & Fuzzy Logic", 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 2015.

[2] H. Waghmare, R. Kokare, Y. Dandawate, "Detection and Classification of Diseases of Grape Plant Using Opposite Colour Local Binary Pattern Feature and Machine Learning for Automated Decision Support System", 3rd International Conference on Signal Processing and Integrated Networks (SPIN),2017.

[3] P. B. Padol ,A. Yadav, "SVM classifier based grape leaf disease detection", Conference on Advances in Signal Processing (CASP), 2016.

[4] K. kaur, C. Marwaha, "Analysis of Diseases in Fruits using Image Processing Techniques", International Conference on Trends in Electronics and Informatics ICEI , 2017.

[5]M.Chaudhary R.Chavan, S.Durgawali,A.Ghodeswar, "Smart Agriculture: Detection of Disease in Plants using Image Processing " International Journal of Engineering Research & Technology (IJERT),2018.

[6] Mrunalini R. Badnakhe, Prashant R. Deshmukh, "Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering", International Journal of Advanced Research in

Computer Science and Software Engineering, Volume 2, Issue 3, March 2012.
[7]C. Iorga, V.E. Neagoe: “A Deep CNN Approach with Transfer Learning for Image Recognition”, ECAI 2019 - International Conference – 11th Edition, Electronics, Computers and Artificial Intelligence, 2019.
[8] S. Ghosal , K. Sarkar: “Rice Leaf Diseases Classification Using CNN With Transfer Learning”, Proceedings of 2020 IEEE Calcutta Conference (CALCON), 2020.
[9] A. K. Reyes, J. C. Caicedo, and J. E.Camargo. Fine-tuning deep convolutional networks for plant recognition. CLEF(Working Notes), 1391:467–475, 2015.