iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: VI Month of publication: June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43700

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Plant Disease Detection Using Deep Learning

Shivam Verma, Prashant Kumar Choudhary, Suraj Kumar, Prof. Dr. Reena Gunjan
Department of ECE, MIT ADT University

Abstract: Crop diseases are a major threat to food security, but their rapid identification remains difficult inmany parts of the world due to the lack of thenecessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased andhealthy plant leaves collected under controlled conditions, we train a deep convolutional neuralnetwork to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves anaccuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly availableimage datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale

I. INTRODUCTION

The occurrence of plant diseases has a negativeimpact on agricultural production. If plant diseases are not discovered in time, food insecurity will increase [1]. Early detection is the basis for effective prevention and control of plant diseases, and they play a vital role in the management and decision-making of agricultural production. In recent years, plant disease identification has been a crucial issue.

Disease-infected plants usually show obvious marks or lesions on leaves, stems, flowers, or fruits. Generally, each disease or pest condition presents a unique visible pattern that can be used to uniquely diagnose abnormalities. Usually, the leaves of plants are the primary source for identifying plant diseases, and most of the symptoms of diseasesmay begin to appear on the leaves [2].

In most cases, agricultural and forestry experts are used to identify on-site or farmers identify fruit treediseases and pests based on experience. This method is not only subjective, but also time- consuming, laborious, and inefficient. Farmers with less experience may misjudgment and use drugs blindly during the identification process. Quality and output will also bring environmental pollution, which will cause unnecessary economic losses. To counter these challenges, research into the use of image processing techniques for plant diseaserecognition has become a hot research topic.

II. LITERATURE SURVEY

In the Paper-Deep learning for Image-Based Plant detection" the authors Prasanna Movant et al., has proposed an approach to detect disease in plants by training a convolutional neural network. The CNN model is trained to identify healthy and diseased plants of 14 species. The model achieved an accuracy of 99.35% on test set data. When using the model on images procured from trusted online sources, the model achieves an accuracy of 31.4%, while this is better than a simple model of random selection, a more diverse set of training data can aidto increase the accuracy. Also some other variations of model or neural network training may yield higher accuracy, thus paving path for making plant disease detection easily available to everyone. Kulkarni as etal in the paper ―Applying image processing technique to detect plant diseases" a methodology for early and accurately plant diseases detection, using artificial neural network (ANN) and diverse image processing techniques. As the proposed approach is based on ANN classifier for classification and Gabor filter for feature extraction, it gives better results with a recognition rate of up to 91%.

In paper ―Plant disease detection using CNN and GAN" by Emaneul Cortes, an approach to detect plant disease using Generative Adversarial networks has been proposed. Backgroundsegmentation is used for ensuring proper feature extraction and output mapping. It is seen that using Gans may hold promise to classify diseases in plants, however segmenting based on background did not improve accuracy.

III. DATASET DESCRIPTION

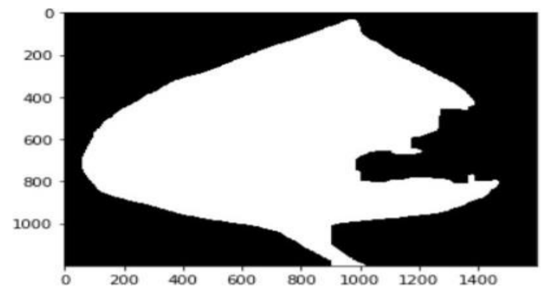
We analyze 54,306 images of plant leaves, which have a spread of 38 class labels assigned to them. Each class label is a crop-disease pair, and we make an attempt to predict the crop-disease pair given just the image of the plant leaf. Figure 1 shows one example each from every crop-disease pair from the PlantVillage dataset. In all theapproaches described in this paper, we resize the images to 256 × 256 pixels, and we perform both the model optimization and predictions on these downscaled images.

Across all our experiments, we use three different versions of the whole PlantVillage dataset. We startwith the PlantVillage dataset as it is, in color; then we experiment with a gray-scaled version of the PlantVillage dataset, and finally we run all the

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

experiments on a version of the PlantVillage dataset where the leaves were segmented, hence removing all the extra background information which might have the potential to introduce some inherent bias in the dataset due to the regularized process of data collection in case of PlantVillage dataset. Segmentation was automated by the means of a script tuned to perform well on our particular dataset. We chose a technique based on a set of masks generated by analysis of the color, lightness and saturation components of different parts of the images in several color spaces (Lab and HSB). Oneof the steps of that processing also allowed us to easily fix color casts, which happened to be very strong in some of the subsets of the dataset, thus removing another potential bias.

This set of experiments was designed to understandif the neural network actually learns the "notion" ofplant diseases, or if it is just learning the inherent biases in the dataset. Figure 2 shows the different versions of the same leaf for a randomly selectedset of leaves.



IV. MEASUREMENT OF PERFORMANCE

To get a sense of how our approaches will perform on new unseen data, and also to keep a track of if any of our approaches are overfitting, we run all our experiments across a whole range of train-test set splits, namely 80–20 (80% of the whole dataset used for training, and 20% for testing), 60–40 (60%of the whole dataset used for training, and 40% for testing), 50–50 (50% of the whole dataset used for training, and 50% for testing), 40–60 (40% of the whole dataset used for training, and 60% for testing) and finally 20–80 (20% of the whole dataset used for training, and 80% for testing). It must be noted that in many cases, the PlantVillage dataset has multiple images of the same leaf (taken from different orientations), and we have the mappings of such cases for 41,112 images out ofthe 54,306 images; and during all these test-train splits, we make sure all the images of the same leaf goes either in the training set or the testing set. Further, for every experiment, we compute the mean precision, mean recall, mean F_1 score, along with the overall accuracy over the whole period of training at regular intervals (at the end of every epoch). We use the final mean F_1 score for the comparison of results across all of the different experimental configurations.

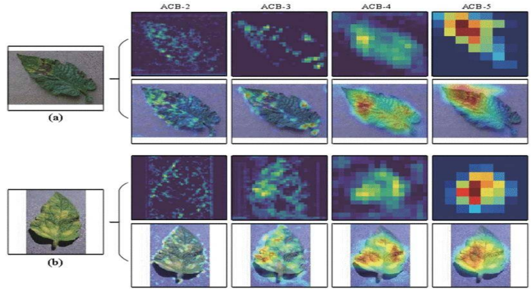
V. VISUALIZATION TECHNIQUE

In recent years, the successful application of deep learning technology in plant disease classification provides a new idea for the research of plant disease classification. However, DL classifiers lack interpretability and transparency. The DL classifiers are often considered black boxes without any explanation or details about the classification mechanism. High accuracy is not only necessary for plant disease classification but also needs to be informed how the detection is achieved and which symptoms are present in the plant. Therefore, in recent years, many researchers have devoted themselves to the study of visualization techniques such as the introduction of visual heat maps and salient maps to better understand the identification of plant diseases. Among them, the works of [35] and [36] are crucial to understanding howCNN recognizes disease from images.

For example, Brahimi *et al.* [35] introduced saliency maps to visualize the symptoms of plant diseases. Mohanty *et al.* [10] used AlexNet and GoogLeNet architectures, through the precision (P), recall (R), F1 score, and the overall accuracy to

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

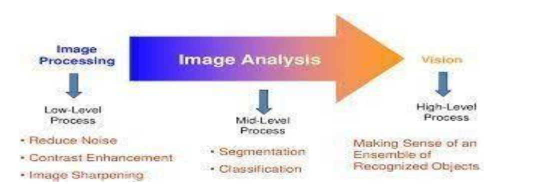
evaluate the performance of the models on the PlantVillage. Used the three scenarios (color gray and segmentation) to assess the performance of the 2 CNN famous architectures, and come to the conclusion that GoogLeNet outperformed AlexNet,the first layer of the visual results clearly showed the disease spots also. In Cruz *et al.* [37], the improved LeNet model was used to detect olive plant diseases, that is, segmentation and edge maps were used to identify plant diseases. Brahimi *et al.* [38] proposed a new visualization method, that is, a new DL model teacher/student network was introduced to identify the spots

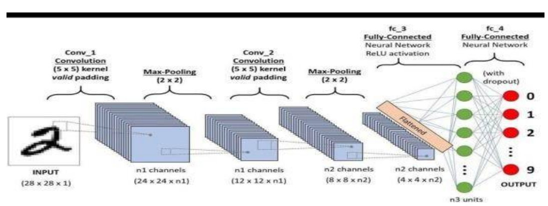


of plant diseases, compared with the existing plant disease treatment methods, the new method obtained a clearer visualization effect.

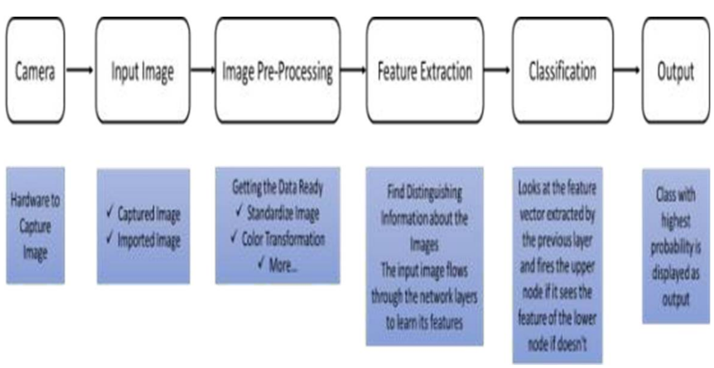
VI. CONVOLUTIONAL NEURAL NETWORK

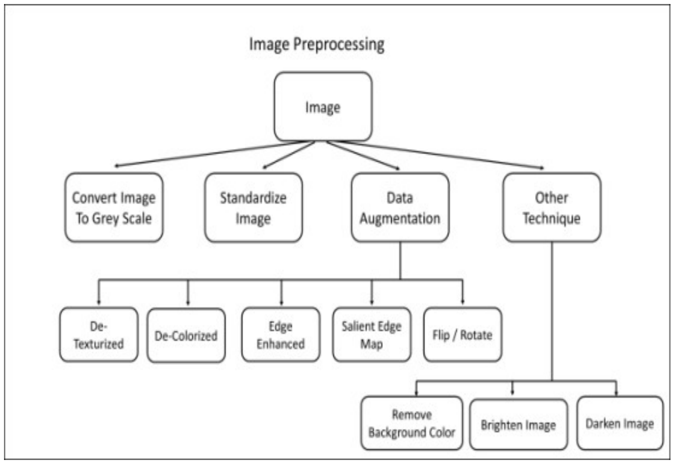
A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural network that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be.
The human brain processes a huge amount of information the second we see an image. Each neuron works in its own receptive field and is connected to other neurons in a way that they coverthe entire visual field. Just as each neuron responds to stimuli only in the restricted region of the visual field called the receptive field in the biological vision system, each neuron in a CNN processesdata only in its receptive field as well. The layers are arranged in such a way so that they detect simpler patterns first (lines, curves, etc.) and more complex patterns (faces, objects, etc.) further along. By using a CNN, one can enable sight to computers.





VII. BLOCK DIAGRAM





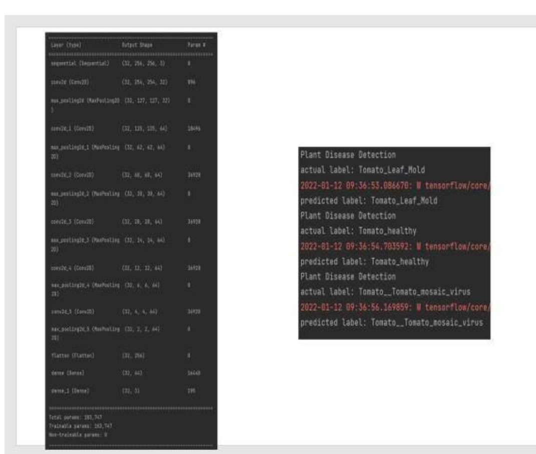
VIII. DISCUSSION

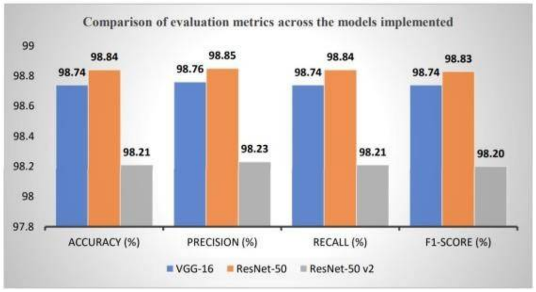
Using the deep convolutional neural network architecture, we trained a model on images of plant leaves with the goal of classifying both crop species and the presence and identity of disease on images that the model had not seen before. Within the Plant Village data set of 54,306 imagescontaining 38 classes of 14 crop species and 26 diseases (or absence thereof), this goal has been achieved as demonstrated by the top accuracy of 99.35%. Thus, without any feature engineering, the model correctly classifies crop and disease from 38 possible classes in 993 out of 1000 images.Importantly, while the training of the model takes a lot of time (multiple hours on a high performance GPU cluster computer), the classification itself is very fast (less than a second on a CPU), and can thus easily be implemented on a smartphone. This presents a clear path toward smartphone-assistedcrop disease diagnosis on a massive global scale.

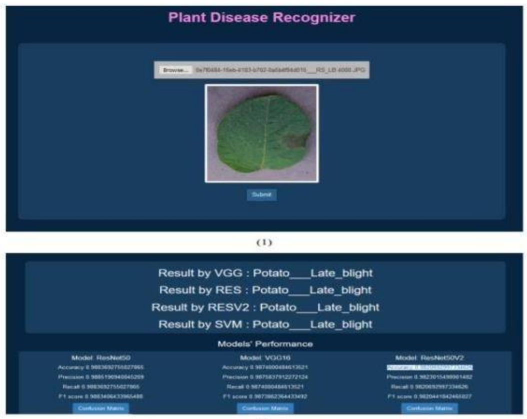
However, there are a number of limitations at the current stage that need to be addressed in future work. First, when tested on a set of images taken under conditions different from the images used for training, the model's accuracy is reduced substantially, to just above 31%. It's important to note that this accuracy is much higher than the one based on random selection of 38 classes (2.6%), butnevertheless, a more diverse set of training data is needed to improve the accuracy. Our current resultsindicate that more (and more variable) data alone will be sufficient to substantially increase the accuracy, and corresponding data collection efforts are underway.

The second limitation is that we are currently constrained to the classification of single leaves,facing up, on a homogeneous background. While these are straightforward conditions, a real world application should be able to classify images of a disease as it presents itself directly on the plant. Indeed, many diseases don't present themselves on the upper side of leaves only (or at all), but on many different parts of the plant. Thus, new image collection efforts should try to obtain images from many different perspectives, and ideally fromsettings that are as realistic as possible.

IX. RESULT DETAILS







X. FUTURE ENHANCEMENT

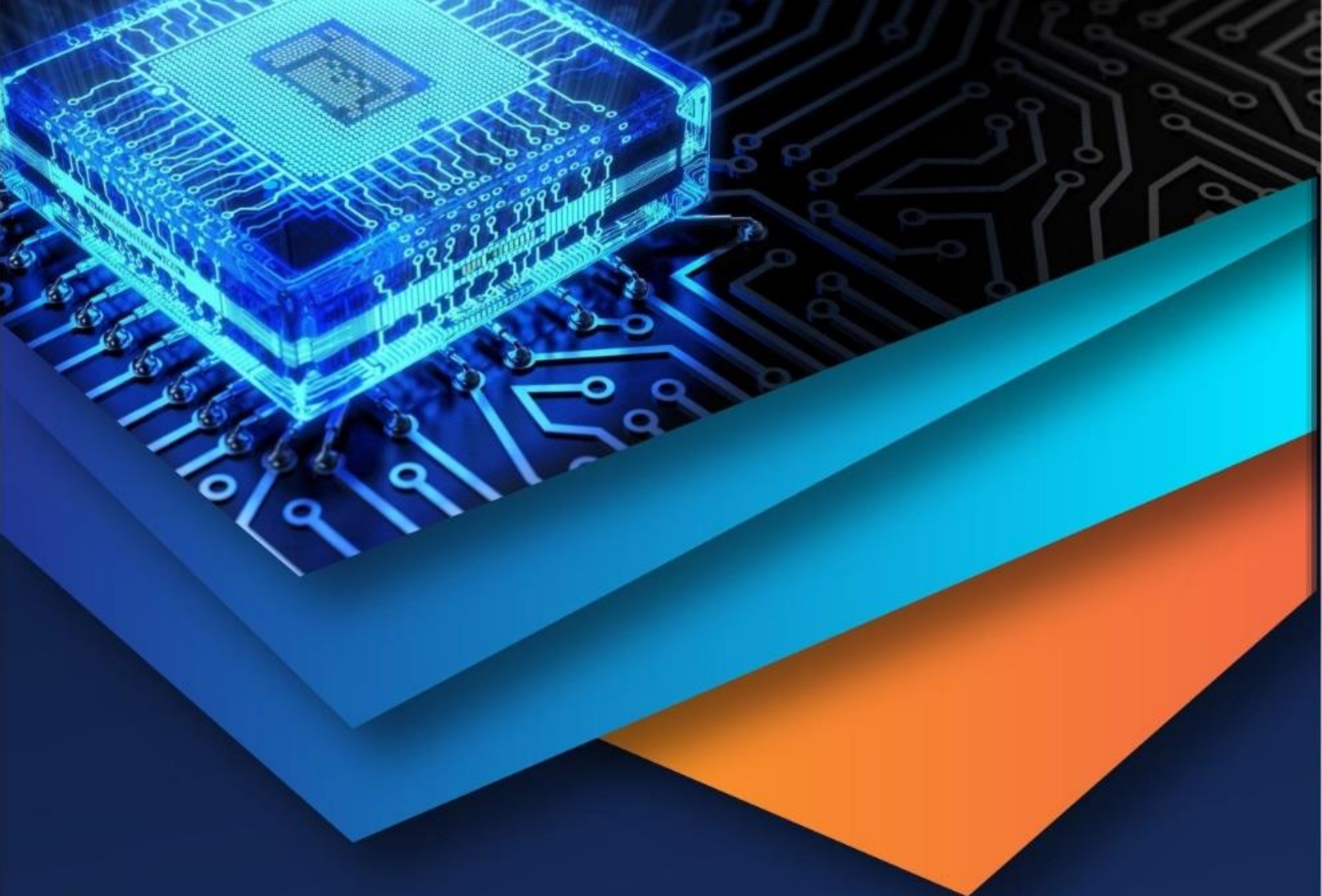
The disease detection system can be integrated in cloud system for efficient result processing. Integration of automated disease detection system with sensors to measure soil.

XI. CONCLUSION

The Deep learning algorithm used in proposed work is SVM. SVM gave good result in when the detection categories were less. As the no of disease categories increased it failed to achieve the accuracy. Transfer learning is the current effective research for obtaining the better performance of the models with a minimal and faster training phase. It proved very true with the proposed framework. The proposed framework able to attain better accuracy with all the three models such as VGG-16, ResNet-50, and ResNet-50v2, yet ResNet-50 based transfer learning model a bit more efficient when compared to the other models. The proposed framework efficient with themulticlass classification of various diseases alongwith healthy leaves that include crops of pepper, potato, and tomato.

REFERENCES

[1] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 18: (1) Input Interface (2) Interface output.
[2] GSMA Intelligence (2016). The Mobile Economy- Africa 2016. London: GSMA.
[3] Raza, S.-A., Prince, G., Clarkson, J. P., Rajpoot, N. M., et al. (2018). Automatic detection of diseased tomato plants using thermal and stereo visible light images. PLoS ONE 10:e0123262. doi: 10.1371/journal.pone.0123262
[4] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2019). ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252. doi: 10.1007/s11263-015-0816y
[5] Lowe, D. G. (2019). Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis.60,91110.doi:10.1023/B:VISI.0000029664.99615.9



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY