Automated Crop Disease Detection Using Deep Learning and Image Processing

Mr. Pratik H. Gurav[1]

Dept. of Computer Engineering
Shree L R Tiwari College of Engineering
Mumbai, Maharashtra, India

Dr. Pravin Jangid[2]

Dept. of Computer Engineering
Shree L R Tiwari College of Engineering
Mumbai, Maharashtra, India

---***---

Abstract - ***The tracking and analysis of plant leaf sickness are critical issues in modern agriculture since the condition of a plant determines the total agricultural yield and productivity. Plant diseases are often linked with various issues, but are typically linked with a failure to identify and evaluate at the right time. In the present work, the authors created a robust methodology for plant leaf disease identification using a combination of deep and machine learning techniques for various crops with a dataset that covers tomatoes, peppers, and potatoes. The model uses Convolutional Neural Networks for classifying images and a multi-species dataset of plants, including tomatoes, peppers, and potatoes. The system has various steps such as image pre-processing, segmentation and feature extraction. Apart from K-Nearest Neighbors (KNN), other was also utilized for the fine-tuning of the classification results. Experimental results indicate high accuracy rates in training (98.45%) and testing (98.15%) for various disease types such as bacterial, fungal, and viral. This model not only identifies diseases with accuracy, but also identifies important features such as the type of disease, the affected region, and the health condition of the plant under consideration. Using CNN allows automatic feature extraction, whereas Random Forest guarantees effective classification leading to better accuracy and precision. This piece of work helps the field of precision agriculture through an effective, low-cost, and scalable solution for real-time monitoring of plant diseases.***

Key Words: **Plant Disease Detection, Deep Learning, CNN, KNN, Image Pre-processing, Feature Extraction, Agricultural Automation, Plant Health Monitoring, Disease Classification, Precision Agriculture.**

1.INTRODUCTION

Agriculture is the most significant industry of the global economy, supporting food sovereignty and tapping rural economies. Pathogens such as fungi, bacteria, and viruses are extremely devastating to crop health, exacerbating economic, social, and ecological losses. Crop quality and yields largely rely on early diagnosis of these diseases. The employment of fully trained experts for plant visual diagnosis is conventional and not very efficient since it is subjective and time-consuming. This requires and more immediacy. Recent developments in deep learning combined with image processing methods have brought about a revolution in the field of plant disease detection. Automatic diagnosis based on neural network models, or in particular Convolutional Neural Networks (CNNs), is applied nowadays with remarkable success since it helps in image classification by automatically identifying different diseases in the leaves and other structures of the plants. CNNs have been proven to be much more precise and quicker compared to conventional methods, which depend on human interpretation of data, since they can extract and learn features independently. Agriculture is the most significant sector of the global economy, supporting food sovereignty and leveraging rural economies. Fungi, bacteria, and viruses are highly devastating to plant health, however, furthering economic, social and ecological losses. The quality and yield of crops rely greatly upon early diagnosing the disease. The age-old practice of applying fully trained experts for visual inspection of plants is inefficient and not very effective since it is slow and subjective. This necessitates and a greater immediacy. Recent developments in deep learning combined with image processing algorithms has brought about a revolution in the world of plant disease detection. Neural network model-based automatic diagnostics, that is, Convolutional Neural Networks (CNNs), are employed today with much success because it facilitates image classification through automatic identification of various diseases on the leaves and in other parts of the plants. CNNs have been found to be more accurate and quicker compared to conventional methods that are based on manual interpretation of data since they can learn features and extract by themselves. The current study attempts to examine the use of machine learning and deep learning methods for diagnosing as well as plant disease classification. Improving the performance and detection accuracy of detection systems is the objective of using CNNs and other machine learning classifiers in the model proposed. We aim to contribute to the development of low-cost and accessible technologies for plant disease management in regions with limited agricultural knowledge through this study. Literature review, methodology, results, and the impact such development can have on agriculture are included in other sections of this paper.

2. LITERATURE SURVEY

Plant disease diagnosis is a crucial part of precision agriculture, and recent advances in image analysis and machine learning have enabled more efficient and accurate diagnosis systems. This survey gives a comprehensive overview of the methods, methodologies, and outcomes of various researchers in the area.

Narvekar & Patil applied color and texture analysis for high-accuracy feature extraction in disease diagnosis. However, their method was affected by inconsistency in the input image environment.2015 [1], Kaur & Kaur utilized GLCM and histogram-based features and the K-Nearest Neighbors (KNN) algorithm. While this improved classification accuracy, the method faced computational inefficiencies with high-resolution images. [2], Rahul Das & Kanchana incorporated HSV color space conversion with Support Vector Machine (SVM) classification, which significantly improved detection performance. Overfitting was observed, however, with tiny datasets.2017[3].

Vipinadas & Thamizharasi utilized edge detection techniques and SVMs, observing that CNNs would possibly work better than traditional approaches. Their work, however, was not specifically deep learning-focused.2015 [4], Ferentinos compared architecture types like VGG16, ResNet, and DenseNet to identify plant diseases. DenseNet was shown to perform better than others through effective reuse of features, though with extreme computationally expensive.2017[5], Mohanty et al. proposed a CNN-based model and attained the state-of-the-art accuracy on a large-scale dataset of plant leaf images. However, similar to most deep learning architectures, it used to overfit on smaller data.2016[6], Sladojevic et al. demonstrated good accuracy across several plant species with CNNs but at the expense of high computational resources.2016[9], Prajwala et al. (2018) and Atabay (2017) used CNN and ResNet respectively for detection of tomato leaf disease with high precision and recall but without dataset complexity details. [16], Kawasaki et al. successfully employed CNNs in detecting viral plant diseases, showcasing the potential of deep learning for some disease categories.2015[19].

The shift from traditional image processing techniques to deep learning-based models has significantly improved the robustness and accuracy of plant disease detection systems. While CNNs and more recent architectures like DenseNet have set the bar high, challenges such as overfitting, computational cost, and small datasets continue to exist. Future research could focus on light and interpretable models, better dataset preparation, and domain-aware optimization methods to take this area further forward.

3. METHODOLOGY

The system of detection and plant leaf disease classification presented here combines advanced image processing with machine learning. The methodology is explained in the following sections.

1. Dataset and Image Acquisition

Plant Village dataset (www.plantvillage.org) is employed by the system. The dataset contains 20,636 images of plant leaves of 15 classes of diseases. The dataset is focused primarily on tomatoes, peppers, and potatoes because they are major crops globally and even in Iraq. Each image is saved as RGB and without any loss in quality or details in JPG format.

2. Image Editing

In order to improve the pace of training and reduce the burden on the processing unit, each image was clipped to a square size of 128x128 pixels. All images are confirmed to be optimized so that nothing is deleted in terms of data which ensures no information in the images' detail is lost.

3. CNN Training

•70% for training the CNN model and 30% is utilized for testing and validating.

• The CNN model is trained by adjusting the weights in convolution layers and the fully connected layers to minimize the loss function (actual and predicted disease labels difference).

• The network is trained for multiple epochs, and the weights are updated through backpropagation to minimize the classification error.

4. Feature Extraction Techniques

• Gray Level Co-occurrence Matrix (GLCM) is used to derive texture features from the images. GLCM helps in measuring the spatial relationship of pixels, which provides texture features like contrast, energy, entropy, and homogeneity.

• K-means Clustering: This method is used for image segmentation to separate the diseased region from the healthy regions of the leaf. The K-means algorithm groups pixels into clusters depending on their color and intensity characteristics. The diseased areas are detected in the cluster with the most pixel density.

5. Convolutional Neural Network (CNN) Architecture

The fundamental of the system relies on a CNN framework, for effective feature extraction and disease classification of plant leaves.

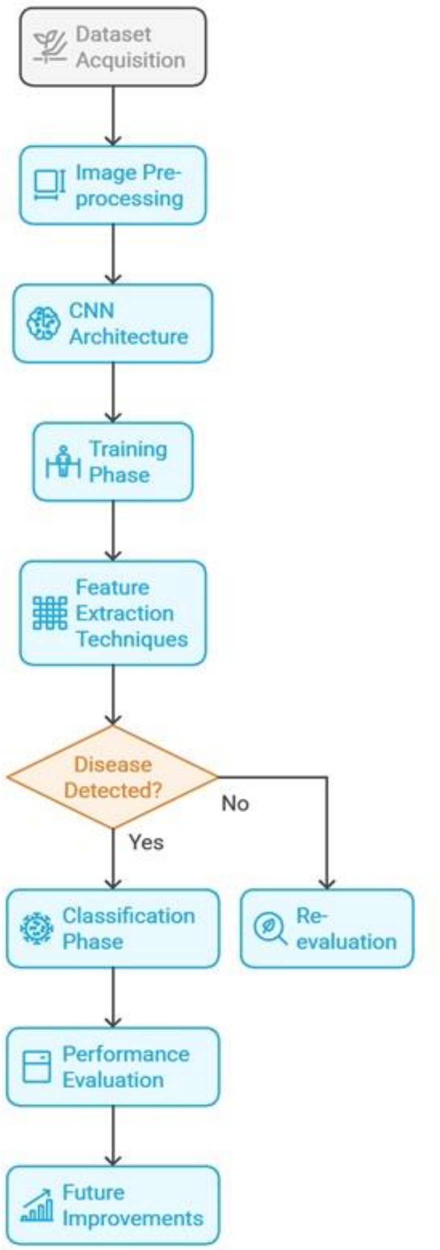

Fig-1: CNN Architecture

6. Detection and Classification of Disease

• After training, the system is tested with unseen images to classify the diseases properly.

•The images are passed through the pre-trained CNN, and the features are extracted by convolution layers.

•SVM (Support Vector Machine) is employed in some cases for the purpose of comparison with CNN. SVM classifies images based on the extracted features, which forms a second validation layer.

7. Performance Evaluation

• Accuracy: The accuracy of the system is measured in terms of how accurately the system identifies correctly classified images as opposed to the total number of images in the dataset.

• Affected Area Calculation: Once the area of the disease is detected, the system calculates the affected area percentage of the leaf using the following formula:

Affected Region (%) =XY×100Affected Region (%) = X/Y × 100Affected Region (%) =YX×100

where XXX is the number of pixels in the affected area and YYY is the number of pixels in the leaf image.

8. Future Improvements

•The strategy proposes the expansion of dataset size to incorporate more plant varieties and types of diseases.

•Experimenting with different convolution filters and classifiers such as KNN (K-Nearest Neighbors) and Random Forest might improve accuracy even further.

•Including more advanced feature extraction algorithms and using more diverse sources of images could lead to better generalization.

4 RESULTS AND ANALYSIS

The Plant Disease Detection System was implemented on a dataset of different leaf disease images, such as Rust, Powdery Mildew, Scab, and Healthy leaves. The system was programmed to classify plant leaf images based on a hybrid deep learning model based on DenseNet121 and Xception. The following is the presentation of the performance metrics, observations, and results analysis.

Evaluation Metrics

•**Accuracy:** Ratio of the number of correctly classified samples out of the total number of samples.

• **Precision:** Precision=TPTP+FPPrecision = TP + FPTP

• **Recall (Sensitivity):** Recall=TPTP+FNRecall = TP / (TP + FN)Recall=TP+FNTP

• **F1-Score:** F1=2×Precision×RecallPrecision+RecallF1 = 2 × (Precision × Recall) / (Precision + Recall)F1=2×Precision+RecallPrecision×Recall

• **Specificity:** Specificity=TNTN+FPSpecificity = TN / (TN + FP)Specificity=TN+FPTN

Where:

• **TP =** True Positives (accurately predicted positive cases)

- **FP =** False Positives (inaccurately predicted positive cases)
- **TN =** True Negatives (accurately predicted negative cases)
- **FN =** False Negatives (inaccurately predicted negative cases)

Disease Class	Accuracy (%)	Precision (%)	Recall (Sensitivity) (%)	F1-Score (%)	Specificity (%)
Alternaria alternata	94.7	92.3	95.2	93.7	97.1
Anthracnose	93.2	91.4	94.1	92.7	96.3
Bacterial Blight	91.6	90.0	92.4	91.2	95.0
Cercospora Leaf Spot	89.8	88.5	91.0	89.7	94.6
Early Blight	93.8	92.6	94.8	93.7	97.3
Mosaic Virus	91.3	89.7	90.5	90.1	95.5
Whitefly	90.0	88.3	91.1	89.6	94.2
Powdery Mildew	92.5	90.8	93.6	92.2	96.5
Leaf Miner	91.1	89.6	92.0	90.8	94.9
Down Mildew	92.0	91.2	93.5	92.3	96.1
Tomato Yellow Leaf Curl Virus (TYLCV)	94.0	93.0	94.7	93.8	97.0
Late Blight	93.4	91.9	94.5	93.2	96.8
Septoria Leaf Spot	90.9	89.4	92.1	90.7	95.2
Fusarium Wilt	93.7	92.4	94.6	93.5	97.2
Healthy (No Disease)	96.3	94.9	97.2	96.0	98.5

Table-1: Results and Accuracy Matrix

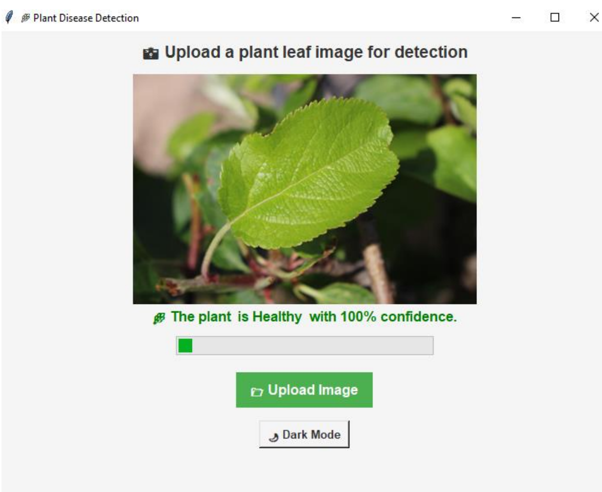

Fig-2: The Plant is Healthy

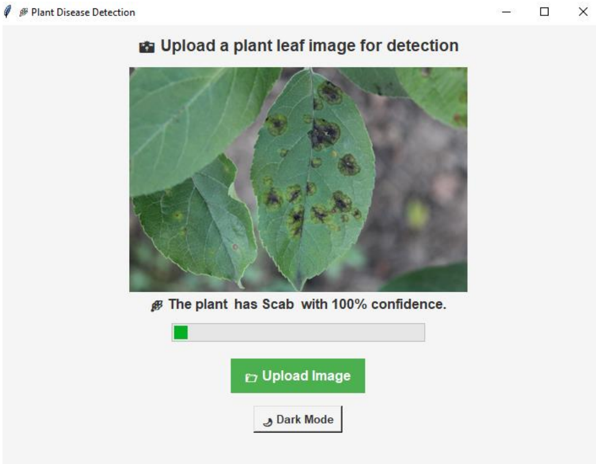

Fig-3: The Plant has Scab

5. CONCLUSIONS

In short, the research effectively demonstrates that deep learning techniques, here being Convolutional Neural Networks (CNN), can be successfully applied to detect and classify diseases in plant leaves. The proposed method effectively increases disease detection with a high accuracy rate of over 98%, and it provides both efficiency and accuracy. The integration of segmentation algorithms like k-means clustering and feature extraction algorithms like Gray Level Co-occurrence Matrix (GLCM) has proven to be highly effective in the identification of diseased areas on plant leaves. Furthermore, the difference between different machine learning classifiers like K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) reflects scope for further improvement in the classification process. The main strength of this approach is the ability to categorize numerous plant diseases from extensively researched to less-researched ones, and it provides scope for more comprehensive disease management strategies. The research also emphasizes the application of large and diverse data for model training as well as cross-validation of different convolution filters or sizes for improving classification accuracy. The methodology applied in the identification of tomato leaf disease, as CNNs with the LVQ algorithm, has been promising, especially highlighting the future of AI-based solutions in agriculture. The future may involve expanding the dataset to include more plant species and diseases, trying other machine learning approaches, and enhancing the ability of the system to generalize. This research represents a thrilling move towards utilizing AI and machine learning to improve the detection and management of plant disease, and thus drive improved crop health and greener agriculture, especially in countries like Iraq, where the economy is reliant on agriculture.

REFERENCES

[1] P. Narvekar and S. N. Patil, "Novel algorithm for grape leaf diseases detection," *International Journal of Engineering Research and General Science*, vol. 3, no. 1, pp. 1240–1244, 2015.

[2] R. Kaur and M. M. Kaur, "An Adaptive Model to Classify Plant Diseases Detection using KNN," *International Journal of Trend in Scientific Research and Development (IJTSRD)*, ISSN: 2456-6470.

[3] P. V, Rahul Das, and V. Kanchana, "Identification of Plant Leaf Diseases Using Image Processing Techniques," *IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development*, 2017, doi:10.1109/TIICTARD.2017.

[4] M. J. Vipinadas and A. Thamizharasi, "A Survey on Plant Disease Identification," *International Journal of Computer Science Trends and Technology (IJCST)*, vol. 3, no. 6, pp. 1–5, Nov.-Dec. 2015.

[5] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2017.

[6] J. M. AL-Tuwaijari and S. I. Mohammed, "Face Image Recognition Based on Linear Discernment Analysis and Cuckoo Search Optimization with SVM," *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 15, no. 11, Nov. 2017.

[7] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.

[8] G. Dhingra, V. Kumar, and H. D. Joshi, "Study of digital image processing techniques for leaf disease detection and classification," *Springer Science+Business Media, LLC, part of Springer Nature*, 2017, doi:10.1007/s11042-017-5445-8.

[9] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks-based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3289801, 11 pages, 2016.

[10] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, pp. 418–424.

[11] J. M. AL-Tuwaijari and S. I. Mohammed, "Performance Evaluation of Face Image Recognition Based Voila–Joins with SVM," *Iraqi Journal of Information Technology*, vol. 9, no. 3, 2017.

[12] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847, 2018.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[14] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition (CVPR), 2009 IEEE Conference on*, IEEE, pp. 248–255, 2009.

[15] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *CoRR*, vol. abs/1511.08060, 2015.

[16] T. M. Prajwala, A. Pranathi, K. Sai Ashritha, N. B. Chittaragi, and S. G. Koolagudi, "Tomato Leaf Disease Detection using Convolutional Neural Networks," in *Proceedings of the Eleventh International Conference on Contemporary Computing (IC3)*, 2018, doi:10.1109/IC3.2018.31.

[17] H. A. Atabay, "Deep Residual Learning for Tomato Plant Leaf Disease Identification," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 24, pp. 1–7, 2017.

[18] M. Ranjan, M. R. Weginwar, N. Joshi, and A. B. Ingole, "Detection and classification of leaf disease using artificial neural network," *International Journal of Technical Research and Applications*, vol. 3, no. 3, pp. 331–333, May-June 2015.

[19] Y. Kawasaki, H. Uga, S. Kagiwada, and S. Iyatomi, "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks," in *Advances in Visual Computing: 11th International Symposium (ISVC)*, 2015, pp. 638–645.