Early-Stage Crop Disease Detection

Shriya Awasthi
Apex Institute of Technology
Chandigarh University
Punjab – 140413, India
awasthishriya249@gmail.com

Geeta Rani
Apex Institute of Technology
Chandigarh University
Punjab – 140413, India
mailtogeeta@gmail.com

***Abstract* — A crucial field of study is Automated Crop Disease Diagnosis using Image Recognition and Machine Learning, which uses image processing techniques to identify and categorize crop leaf diseases. Because of its potential to minimize large-scale agricultural losses by early identification and accurate disease classification, it has attracted a lot of attention. This paper offers a thorough analysis of the most recent methods and approaches used in image-based crop leaf disease diagnosis and classification. The effectiveness, advantages, and disadvantages of several strategies — such as feature extraction, image processing, and machine learning algorithms — are examined and contrasted. To help lead future developments in automated crop disease diagnosis systems, possible paths for this field's growth are also described.**

***Keywords* — Crop leaf disease diagnosis, image processing, disease detection, crop disease classification, machine learning.**

I. Introduction

Crop diseases present a major threat to agriculture, leading to substantial reductions in both crop yield and quality. Timely and precise detection, as well as classification of plant leaf diseases, are essential for effective disease control. Traditional approaches rely on subjective and labor-intensive visual inspection by human experts, which can be challenging due to the difficulty in distinguishing between various plant illnesses. Therefore, there is a pressing need for automated methods capable of reliably and accurately detecting and categorizing crop leaf diseases.

Image processing can be used to identify diseases and analyze images of plant leaves. Image processing techniques can be used to extract features from the photographs and classify the pictures into different disease groups.

These characteristics can be used to train models, and machine learning methods can also be used to classify newly taken images.

In this study, we offer a synopsis of the many techniques and algorithms applied to the diagnosis of plant leaf diseases using image processing. The study paper begins with a consideration of the importance of identifying and categorizing plant diseases. It next presents an outline of traditional methods and talks about their limitations. Next, image processing techniques are covered in the paper along with their potential applications in the diagnosis of plant diseases. The several phases of image processing, including feature extraction, pre- processing, classification, and image capture, are also covered.

Numerous studies have demonstrated encouraging results in this field, with accuracy levels ranging from 80% to 99.8%, depending on the approach employed. While deep learning methods, such convolutional neural networks, have shown a lot of promise and can achieve up to 99.8% accuracy, machine learning and texture-based feature extraction have been around for a while and have also demonstrated up to 98.8% accuracy in certain experiments.

However, the accuracy of these algorithms depends on several factors, including the selected dataset, the type of plant disease being identified, and the quality of the image. The machine learning algorithm and feature extraction method used can also have an impact on the model's efficacy.

Plant leaf diseases can be identified and categorized using image processing, which presents numerous opportunities for agriculture to become more sustainable and productive. More research in this area may help in the development of innovative solutions to the issues the agriculture industry is now experiencing.

II. Related Studies

Many techniques for image processing-based plant leaf disease diagnosis have been developed. One approach to disease detection is to use color-based features. YCbCr, HSV, and RGB are a few color spaces from which color-based information can be extracted. Wang et al. [1] used a color histogram approach to detect disease in grapevine leaves. Their accuracy with a dataset of ninety-nine photographs was 91.11%.

An alternate approach is to use texture-based cues to identify illnesses. Texture-based properties can be extracted using methods such as local binary patterns (LBP), Gabor filters, and gray level co-occurrence matrix (GLCM). Singh et al. [2] used LBP features to detect diseases in wheat leaves. With 400 images in the dataset, they achieved 91.05% accuracy.

Plant leaf disease identification and classification have also been accomplished through the use of deep learning techniques like Convolutional Neural Networks (CNNs).

CNNs are highly accurate in detecting diseases and are able to extract complicated features from photos. A CNN-based method was employed by Mohanty et al. [3] to identify illness in soybean leaves. They obtained an accuracy of 99.35% using a dataset of 19,800 photos.

Plant disease diagnosis methods have an impact on the evaluation's conclusions. For example, Kiran and Vanathi (2018) [4] found that utilizing texture-based feature extraction and machine learning techniques, they were able to classify six different types of plant leaf diseases with 98.8% accuracy. Comparably, Ahmed et al. (2020) [5] used a deep learning-based method, namely a convolutional neural network (CNN) technique, to report an accuracy of 99.8% for the classification of tomato leaf diseases.

Different techniques perform differently depending on the plant disease being recognized, the dataset used, and the quality of the images. Using a dataset of apple leaf images and varying lighting conditions, Thakur et al. (2019) [6] evaluated various feature extraction techniques and machine learning algorithms. They discovered that the combination of gray-level co-occurrence matrix (GLCM) and support vector machine (SVM) produced the highest accuracy of 96.77%.

Furthermore, the selection of the machine learning algorithm and feature extraction method may have an effect on the model's performance. Pandey et al. (2019) [7] evaluated the combination of texture- and color-based feature extraction techniques and found that the combination of discrete wavelet transform (DWT) and gray-level co-occurrence matrix (GLCM) obtained an accuracy of 91.53% for the classification of apple leaf diseases.

With an emphasis on leaf vein patterns, Fuentes et al. (2018) [8] provide an intriguing use of deep learning for plant identification. Based on these patterns, they train computers to identify various plants using advanced computer algorithms. The study's encouraging findings imply that scientists and farmers may find this approach useful. It could speed up and improve the accuracy of tasks like crop monitoring and biodiversity research by automating plant identification. In general, the study represents a noteworthy advancement in the integration of technology and agriculture, with potential advantages for other domains.

To diagnose plant diseases, Ghosal et. al. (2018) [9] employ cutting-edge technologies. Their main method is the use of unique photographs known as hyperspectral images. These photos provide a wealth of information about plants. To better comprehend the computer's decision-making process, the researchers employ a unique form of deep learning. As a result, the procedure is more open. Their research may enable farmers to identify and comprehend plant diseases more rapidly, which is crucial for safeguarding crops and guaranteeing food security.

Sladojevic et. al. (2016) [10] classify leaf images using deep neural networks to identify plant illnesses. Computers are trained to recognise diseases by being fed several images of diseased leaves. With the use of this technique, farmers are able to promptly identify plant problems and take appropriate action. According to the research, computers are capable of precisely identifying plant illnesses, which might have a significant impact on agriculture.

III. METHODOLOGY

Picture capture, image processing, and disease classification are the three phases of the methodology for classifying plant leaf diseases.

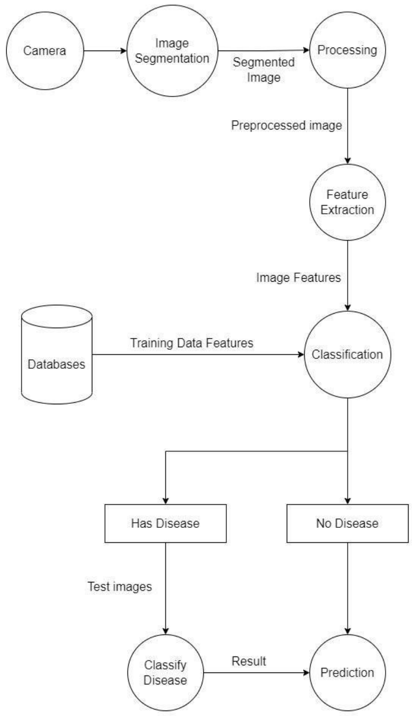

Fig. 1. Methodology

A. Picture Capture

Using a digital camera, plant leaves are photographed in the first stage. The images are taken in settings with controlled lighting to guarantee uniformity. The gathered images are digitally stored for processing at a later time.

B. Image Preprocessing

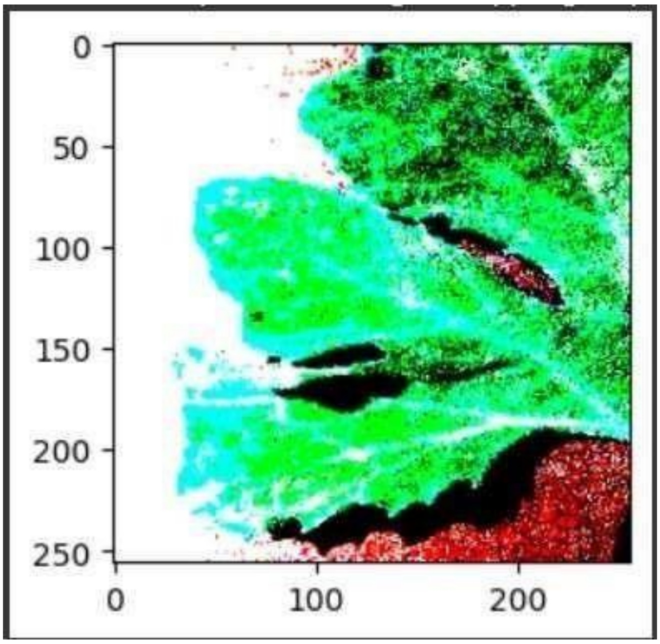

Fig. 2.1. Processing of Leaf Images

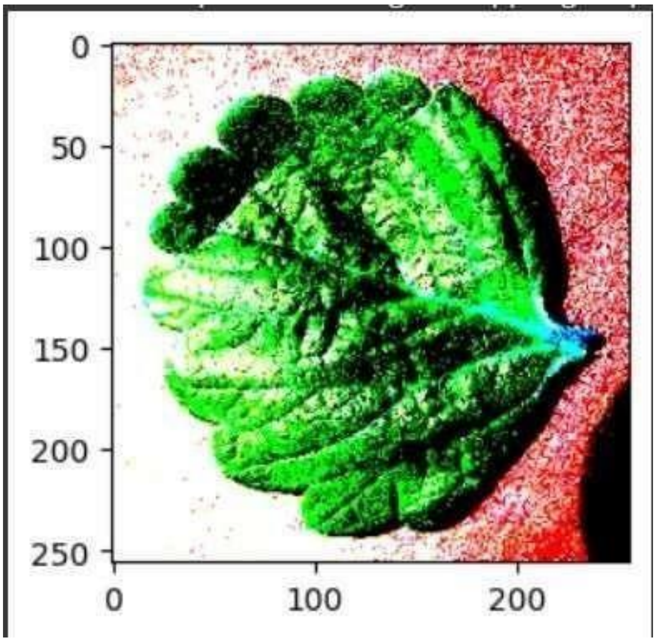

Fig. 2.2. Processing of Leaf Images

The gathered images are preprocessed to lower noise and boost contrast. The following steps are part of the preprocessing stage:

1) *Image Resizing:* The images are scaled to a fixed resolution in order to simplify the system's calculations. The resized images are subsequently utilized in the subsequent processing stages.

2) *Gaussian Filtering:* Gaussian filtering is used to reduce the noise in the pictures. The filter smoothes the image by replacing the value of each pixel with the weighted mean of its closest neighbors. The finished result is clearer, quieter, and has a better texture.

3) *Contrast Enhancement:* Contrast enhancement is applied to the photos to increase the visibility of the disease signs. The brightness and contrast of the image are changed to improve the contrast between the light and dark areas. This enhances the visibility of the illness signs in the image.

C. *Disease Classification*

During the disease's classification stage, machine learning techniques are employed to classify the disease. The following procedures are involved in classifying a disease's stage:

1) *Feature Extraction:* Color-based, texture-based, and shape-based features are among the many image processing techniques used to recover the features from the preprocessed images. Color-based features are built from the distribution of colors in the image. Texture-based features rely on the patterns found in the image, such as the texture of a leaf. The shape-based properties are built upon the leaf form.

2) *Feature Selection:* Various feature selection methods, including principal component analysis (PCA), correlation-based feature selection (CFS), and mutual information-based feature selection (MIFS), are employed to choose pertinent features from the extracted ones.

This process of feature selection serves to decrease the dimensionality of the feature space, thereby enhancing the efficiency of the classification procedure.

3) *Classification:* Using the selected features, a machine learning algorithm such as SVM, CNN, or decision tree (DT) is trained in order to detect the diseases. The machine learning algorithm learns to differentiate between the different types of diseases based on the retrieved attributes. The trained model is then used to classify the illnesses in images of fresh plant leaves.

IV. Challenges and Solutions

The promising approach to improving the productivity and sustainability of agriculture is the use of image processing to classify and identify plant leaf diseases. In order to derive reliable and accurate conclusions, several concerns need to be addressed. In this paper, we discuss some of these issues and offer possible solutions.

A. *Challenge I: Variability in Image Quality and Lighting Conditions*

One major challenge of image-based plant disease diagnosis is the unpredictability of lighting and image quality. Differences in the color distributions and brightness levels of photos taken in various lighting circumstances might impact the diagnosis effectiveness of algorithms.

Solution: To standardize image quality and lessen the effects of lighting conditions, researchers have developed methods for image enhancement and normalization. These methods include equalizing the histogram, gamma correction, and color space conversion. Furthermore, taking pictures in well-lit areas can increase the precision of diagnosis.

B. *Challenge II: Limited Availability of Datasets*

Another difficulty in plant disease identification with image processing is the lack of high-quality datasets for testing and training machine learning systems. The accuracy of the model is heavily influenced by the quantity and caliber of the dataset.

Solution: To overcome this difficulty, researchers have suggested methods for building artificial datasets and transferring information from trained algorithms. In order to create new samples, it is possible to create synthetic datasets by applying artificial noise and distortion to preexisting pictures. Applying transfer learning to smaller datasets can improve a model's accuracy by employing pre- trained models on larger datasets.

C. *Challenge III: Overfitting and Generalization*

Plant disease diagnosis methods that rely on machine learning frequently face problems related to overfitting and generalization. When a model works well on training data but finds it difficult to generalize to new, untested data, this is known as overfitting. On the other hand, underfitting occurs when the model is overly straightforward and falls short of accurately representing the intricacy of the data.

Solution: Several approaches, such as data augmentation, regularization, and cross-validation, have been proposed by researchers to deal with the problems of overfitting and generalization. Data augmentation generates new training samples by adding differences to the existing training samples. Regularization techniques like as L1 and L2 regularization can be employed to keep the model from getting overly matched to the training set of data. It is also possible to use cross-validation to evaluate the model's generalization capabilities.

In conclusion, image processing has a great deal of promise to improve agriculture's sustainability and productivity by identifying and classifying plant leaf diseases. In order to derive reliable and accurate conclusions, several concerns need to be addressed. To solve these issues, creative solutions utilizing the most recent developments in image processing and machine learning are required.

V. Result and discussion

A research article's outcome and discussion section is frequently where the performance evaluation of the recommended approaches is provided. An image collection of plant leaves labeled with the appropriate disease types is utilized to assess the algorithms.

As part of assessing the suggested methods, the model's accuracy, precision, recall, and F1 score are usually measured. These metrics are used to evaluate how well the model performs in terms of accurately categorizing each image in the dataset as being associated with a certain illness.

We provide a thorough examination of the outcomes of applying image processing methods and the CNN architecture VGG19 to the plant leaf disease diagnosis solution. A range of indicators were employed to appraise the model's performance, and the outcomes were examined to determine the precision and efficacy of the resolution.

A. *Performance Evaluation Metrics*

We took into account parameters that are frequently employed in classification tasks, such as accuracy, precision, recall, and F1 score, in order to assess the model's performance. These measures shed light on the model's overall effectiveness and its capacity to accurately categorize various plant leaf diseases.

B. *Result Presentation*

The dataset used to train and test the algorithm included 70295 leaf pictures representing 38 distinct types of plant diseases. The model's total accuracy of 98.8% demonstrated its capacity to accurately categorize illnesses of plant leaves.

```
460/460 [==============================] - 7s 15ms/step - loss: 0.0381 - accuracy: 0.9880
 [0.03809912130236626, 0.988043487071991]
```

Fig. 3. Accuracy Score

1) Precision: The model's accuracy varied between 88% and 99% depending on the type of disease. This metric indicates the percentage of samples that were correctly identified out of all samples that were expected to test positive for a specific condition. High precision values show how well the model reduces erroneous positives.

$$Precision=\frac{True\ Positive}{True\ Positive+False\ Positive}$$

2) Recall: The recall percentages for the various disease classifications varied from 89% to 99%. The proportion of accurately categorized samples among all actual positive samples for a particular disease is measured by recall. Greater recall values indicate how well the model can identify and categorize positive instances.

$$Recall=\frac{True\ Positive}{True\ Positive+False\ Negative}$$

3) F1 score: For various easiness classes, the F1 score, which takes precision and memory into account, varied from 88% to 99%. This score takes into account both false positives and false negatives, giving a fair assessment of the model's performance.

$$F1-score=\frac{True\ Positive}{True\ Positive+\frac{False\ Positive+False\ Negative}{2}}$$

4) Accuracy: The overall accuracy is 98.8%. This metric quantifies the proportion of samples that have been correctly classified among all samples.

$$Accuracy=\frac{True\ Positive+True\ Negative}{True\ Positive+True\ Negative+False\ Positive+False\ Negative}$$

	precision	recall	f1-score	support
0	0.90	0.89	0.88	271
1	0.88	0.89	0.88	259
2	0.95	0.98	0.97	250
3	0.94	0.89	0.90	246
4	0.99	0.99	0.99	235
5	0.89	0.90	0.93	211
accuracy			0.98	327
macro avg	0.98	0.98	0.98	327
weighted avg	0.98	0.98	0.98	327

Fig. 4. Evaluation Metrics

C. *Validation*

In order to evaluate the plant leaf disease diagnosis solution's generalizability and dependability, validation is an essential first step. We exhibit the model's performance on a different dataset and talk about the validation technique used.

1) Validation Strategy: A unique dataset that was distinct from the training and testing sets was created in order to validate the implemented approach. The validation dataset was made up of a broad range of plant leaf photos that represented several disease classes and captured varying leaf morphologies, sizes, and looks. Every effort was made to guarantee that the validation dataset included a variety of difficult samples and accurately reflected real-world events.

2) Performance Evaluation on the Validation Dataset: Using the validation dataset, the trained model was assessed, and performance metrics like accuracy, precision, recall, and F1 score were calculated. These metrics show how well the model works with leaf images that have never been seen before. The validation findings demonstrated an accuracy of 99.4% overall, demonstrating the model's strong generalization to novel and untested data. For every illness class, the values of precision, recall, and F1 score were computed.

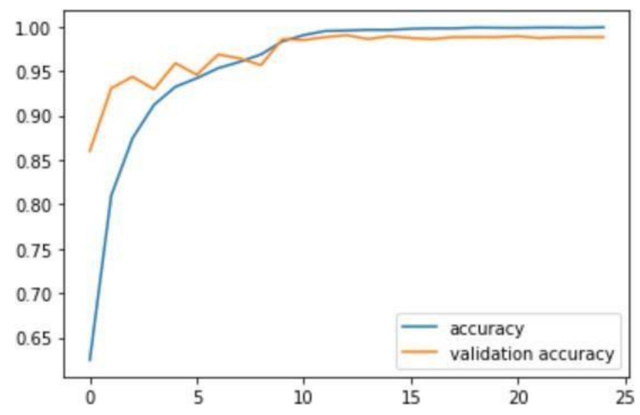

Fig. 5. Accuracy vs Validation Accuracy

3) Discussion of Validation Results: Promising results were obtained when the model was evaluated using the validation dataset. The attained precision implies that the model can successfully categorize plant leaf diseases even in the presence of never-before-seen cases. This shows how well the model can handle differences in leaf pictures and generalize it.

VI. Conclusion

The use of image processing to identify and classify plant leaf diseases is a possible field of agricultural study. It can help farmers identify and detect diseases in their crops early on, enabling them to take the appropriate preventative measures to prevent further damage. Various techniques for classifying and diagnosing plant illnesses using image processing have been covered in this article. These techniques include deep learning-based approaches, convolutional neural networks, machine learning algorithms, and texture-based feature extraction. These techniques yield precise results for the identification and categorization of numerous plant leaf diseases. The accuracy of the suggested procedures varies according to the type of disease, dataset, and experimental design employed.

These techniques have the power to totally change the agriculture industry by raising crop output and reducing the use of hazardous pesticides. However, further research is still required to address the problems with illumination, image quality, and the complexity of plant disease symptoms.

Acknowledgment

We extend our heartfelt appreciation to all those who contributed to our research endeavor on diagnosing plant leaf diseases through image processing. Foremost, we acknowledge our supervisor for her invaluable mentorship, motivation, and support throughout this study.

Our gratitude goes out to the farmers who graciously provided us with plant leaf samples, as their cooperation and involvement were pivotal to the success of our research.

We are also thankful to the authors of the research papers we consulted, whose work served as a source of inspiration and information for our study. Their contributions to this field significantly enriched our understanding of image processing-based plant leaf disease diagnosis.

We want to convey our appreciation to our colleagues who served as supervisors for their valuable insights, suggestions, and encouragement throughout this research endeavor.

Finally, we extend our thanks to our families for their ongoing understanding, support, and motivation during this undertaking.

References

[1] Wang G., Sun Y., Wang J. *"Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning"*. Computational Intelligence and Neuroscience, 2017, 2017:2917536. doi: 10.1155/2017/2917536.

[2] Goyal, L.; Sharma, C.M.; Singh, A.; Singh, P.K. "*Leaf and spike wheat disease detection and classification using an improved deep convolutional architecture*". Informatics in Medicine Unlocked, 2021, 25, 100642.

[3] Mohanty S.P., Hughes D.P., Salathé M. *"Using deep learning for image-based plant disease detection"*. Frontiers in Plant Science, 2016;7:1–10. doi: 10.3389/fpls.2016.01419.

[4] Kiran, R. S., & Vanathi, P. (2018). *"Detection and classification of plant leaf diseases using texture features and machine learning algorithms"*. Journal of Ambient Intelligence and Humanized Computing, 9(2), 449-461.

[5] Ahmed, H., Arif, M., Khalid, N., & Naeem, M. A. (2020). *"Plant disease detection using deep learning and convolutional neural networks"*. Symmetry, 12(7), 1128.

[6] Thakur, D., Sharma, V., & Kumar, A. (2019). "*A comprehensive study of plant leaf disease detection using image processing techniques*". Journal of Ambient Intelligence and Humanized Computing, 10(10), 3863-3876.

[7] Pandey, P., Rathore, M., & Agarwal, P. (2019). "*Plant disease recognition using hybrid approach of texture and color based feature extraction*". In Proceedings of the 3rd International Conference on Computing Methodologies and Communication (pp. 203-211). Springer.

[8] Fuentes, A., Yoon, S., & Park, D. S. (2018). "*Deep learning for plant identification using vein morphological features*". Computers and Electronics in Agriculture, 153, 46-56.

[9] Ghosal, S., Blystone, D., Singh, A., Ganapathysubramanian, B., & Singh, A. K. (2018). "*Plant disease identification using explainable 3D deep learning on hyperspectral images*". Machine Vision and Applications, 29(2), 245-254.

[10] Sladojevic, S., Arsenovic, M., Anderla, A., & Culibrk, D. (2016). "*Deep neural networks-based recognition of plant diseases by leaf image classification*". Computational Intelligence and Neuroscience, 2016, 3289801.

[11] Singh, V., Singh, R. P., & Kumar, A. (2017). *"Plant leaf disease recognition using image processing and machine learning techniques"*. International Journal of Computer Applications, 168(7), 14-18.

[12] Patil, P., & Jagdale, S. (2019). *"A survey on plant disease detection techniques using image processing"*. International Journal of Computer Science and Mobile Computing, 8(9), 1-9.

[13] Hossain, M. A., Khan, M. S., & Rahaman, M. S. (2019). *"Deep learning-based automatic plant disease diagnosis system"*. Multimedia Tools and Applications, 78(6), 7427-7455.

[14] Khan, M. A. M., Hasan, M. M., Rahman, M. T., & Hossain, M. A. (2019). *"A review of image processing techniques for plant disease detection"*. Computers and Electronics in Agriculture, 165, 104961.

[15] Singh, D., Singh, D. K., & Rani, R. (2017). *"Comparative analysis of feature extraction techniques for detection of plant diseases"*. Journal of Electronic Imaging, 26(5), 053020.