WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Image-Based Plant Disease Detection System Using CNN

Dr. Ch. Prem Kumar[1], Ms. Nelluri Renuka Chowdary[2], Ms. Maale Sneha Latha[3]

[1,2,3]Department of IT, MGIT(A), Gandipet, Hyderabad, 500075, Telangana, India.

[1]chpremkumarit@mgit.ac.in, [2]nellurirenukachowdary@gmail.com, [3]maalesnehareddy@gmail.com

ABSTRACT

Plant diseases continue to challenge agricultural productivity, leading to sig- nificant yield losses and economic impact. Early and accurate diagnosis plays a critical role in effective disease management. This project introduces an image- driven solution to classify plant diseases using deep learning models. By leveraging annotated image datasets of diseased plant leaves and fruits, the system identifies the infection and offers two treatment pathways—organic and chemical—tailored to the detected issue. Multiple models were tested, including traditional classifiers and deep learning networks, to benchmark diagnostic performance. Beyond classification, the system recommends prevention and sustainable farming practices, offering a complete decision-support tool for farmers and agronomists.

Keywords: Plant pathology, deep learning, classification, leaf analysis, agriculture tech- nology.

1. Introduction

Agriculture remains vital to global food security and economic stability. However, plant diseases hinder optimal crop yield and threaten the livelihood of farmers. Conventional disease identification methods rely on visual inspection by experts, often constrained by limited accessibility and potential for error.

Recent developments in computer vision and learning algorithms have enabled scal- able solutions for automated disease detection. Deep learning, especially convolution- based architectures, is particularly effective in analyzing visual patterns in plant tissues. This work proposes a system that processes images of leaves and fruits, detects disease symptoms, and suggests treatment options.

The system blends handcrafted and learned features by evaluating multiple algorithms and model structures, including pre-trained models. With a balanced and well-labeled dataset comprising various plant species and disease types, our solution aims to deliver precise results and actionable insights. The ultimate goal is to empower users—especially those in rural or remote areas—with a fast and reliable diagnostic tool accessible via a simple web interface.

2. System Architecture

The architecture comprises the following core components:

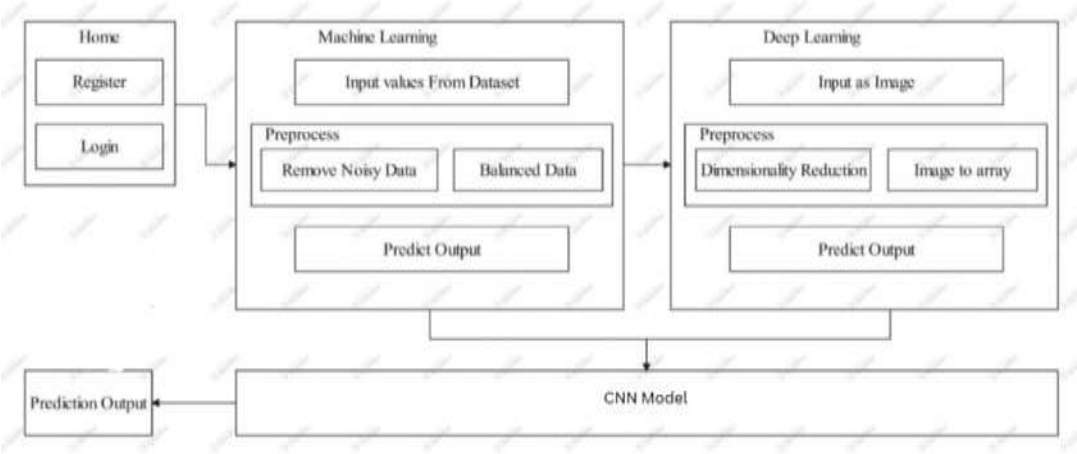

Figure 1: System Architecture of the Plant Disease Detection System

- **Image Input:** Users upload leaf or fruit images via a web interface for real-time analysis.
- **Preprocessing:** Images are resized, normalized, and denoised to enhance quality and model performance.
- **Prediction Module:** The processed image is evaluated by a deep neural model trained to classify disease conditions.
- **Diagnosis Output:** The system returns the predicted class with a confidence score and overlays it on the uploaded image. It further offers treatment sugges- tions—organic and chemical—for each identified condition.

3. Dataset Description and Preprocessing Methodol- ogy

This study utilized a diverse image dataset aggregated from multiple open-source plat- forms, including Kaggle and publicly available academic repositories. The dataset com- prises high-resolution images that capture visual symptoms of plant diseases, with each image labeled according to the disease type and the corresponding crop. In total, the dataset includes samples representing **32 distinct disease categories** spread across a range of agricultural crops such as fruits, vegetables, and grains.

The primary objective of assembling this dataset was to ensure sufficient variation in disease symptoms across crop types, lighting conditions, and leaf orientations to enhance the robustness and generalizability of the machine learning models.

To prepare the images for effective training and evaluation, a comprehensive prepro- cessing pipeline was applied:

- **Uniform Image Resizing:** All input images were resized to a fixed resolution of **150×150 pixels**. This standardization step ensures uniformity across the dataset and allows for efficient training using convolutional neural networks, which require consistent input dimensions.
- **Pixel Intensity Normalization:** To streamline the learning process, the pixel values of each image were normalized by scaling them to a continuous range between **0 and 1**. This is achieved by dividing all pixel values by 255, which is the maximum possible value for 8-bit image data. Normalization facilitates faster convergence during training and minimizes the impact of varying brightness levels across images.
- **Noise Reduction and Image Enhancement:** To improve the clarity of disease features, noise reduction techniques such as Gaussian and median filtering were applied. These filters help in eliminating irrelevant background information and enhancing important patterns like lesion textures, discoloration, and contour details that are critical for accurate classification.
- **Class Distribution Balancing:** As the original dataset contained uneven rep- resentation across disease categories, steps were taken to balance the class distri- bution. Techniques such as data augmentation (e.g., rotation, flipping, zooming) and controlled under-sampling were employed to ensure that each class had an ap- proximately equal number of samples. This is vital to prevent model bias toward overrepresented classes and to improve overall predictive performance.

Following preprocessing, the dataset was partitioned into three subsets to support model development and evaluation:

- **Training Set (70%):** This subset was used to train the machine learning mod- els. It contained the majority of the data and helped the models learn underlying patterns and features associated with different plant diseases.
- **Validation Set (15%):** Used during model training to fine-tune hyperparameters and assess generalization performance. It provides feedback to avoid overfitting and guides adjustments to the model architecture.
- **Testing Set (15%):** This portion was set aside to provide an unbiased evaluation of the final trained model's performance. None of the images in this set were exposed to the model during training or validation.

The dataset split was carried out using stratified sampling to preserve the proportion of each class in all three subsets, thereby maintaining consistency and fairness during model evaluation.

4. Model Training and Development

To effectively classify plant diseases from leaf images, two primary modeling approaches were implemented in this study:

- **Custom Convolutional Neural Network (CNN):** A deep learning model was designed from scratch using a sequential architecture. It included multiple convo- lutional layers for feature extraction, max-pooling layers for spatial dimensionality

reduction, and fully connected dense layers for final classification. The structure was optimized to match the complexity and variability present in the dataset.

- **Transfer Learning with Pre-trained Models:** Advanced convolutional neural networks, including *VGG16* and *ResNet152V2*, were employed using transfer learn- ing. These models were pre-trained on large-scale image datasets like ImageNet and fine-tuned on the current plant disease dataset. This approach allowed for more efficient training and improved accuracy, especially when dealing with limited labeled data.

To improve generalization and prevent model overfitting, several training optimiza- tions were introduced:

- **Data Augmentation:** The training set was enriched using techniques such as image rotation, flipping (horizontal and vertical), and zooming. These transforma- tions helped the model become invariant to image orientation and scale, enhancing its ability to generalize to new data.
- **Regularization Techniques:** *Early stopping* was used to halt training once the validation loss stopped improving, thereby preventing overfitting. Additionally, *dropout layers* were incorporated into the model to randomly deactivate neurons during training, which further improved generalization performance.

Table 1: Comparison of Classification Models

Model	**Accuracy**	**Training Speed**	**Recommended Use Case**
TraditionalClassifiers (e.g., SVM, KNN)	Moderate	Fast	Suitable for binary classification with struc- tured/tabular data
Custom CNN	High	Medium	Optimal for multi-class classification tasks on large image datasets
Pre-trainedCNN (VGG16, ResNet152V2)	Very High	Slow	Best for complex image datasets with limited training data

5. Model Deployment and Evaluation

After training, the final model was deployed using a Flask-based RESTful API, enabling integration with a web-based front-end. Users can upload plant leaf images and receive real-time disease predictions. Image preprocessing and prediction were handled server- side using *OpenCV* for image manipulation and *NumPy* for numerical processing.

5.1 Evaluation Metrics

To assess the classification performance of the model, standard evaluation metrics were utilized. These metrics are based on the outcomes of predictions compared to actual labels, represented using the following terms:

- **True Positive (TP):** The number of correctly predicted positive cases (i.e., the model correctly identifies a diseased leaf).
- **False Positive (FP):** The number of incorrect positive predictions (i.e., the model incorrectly identifies a healthy leaf as diseased).
- **False Negative (FN):** The number of missed positive cases (i.e., the model fails to identify a diseased leaf).
- **True Negative (TN):** The number of correctly predicted negative cases (i.e., the model correctly identifies a healthy leaf).

Using these components, the following metrics were computed:

- **Accuracy:**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy reflects the overall correctness of the model across all classes.

- **Precision:**

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision indicates how many of the predicted diseased leaves were actually diseased.

- **Recall:**

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall shows how many of the actual diseased leaves were correctly identified by the model.

- **F1-Score:**

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{}$$

Precision + Recall

The F1-Score is the harmonic mean of precision and recall, providing a balanced measure when dealing with class imbalance.

- **Detection Time:** The average time taken by the model to classify a single image during inference.

5.2 Model Performance Summary

The final model was trained over 100 epochs, and its performance was summarized as follows:

Table 2: Model Performance Summary

Metric	**VGG16**	**ResNet152V2**	**DenseNet (Final Model)**
Accuracy	81.2%	82.5%	**83.0%**
Precision	80.7%	81.9%	**82.4%**
Inference Time	∼0.50s	∼0.46s	∼0.44s

6. Conclusion

This project successfully demonstrated the application of image processing and deep learning techniques for the detection and classification of plant leaf diseases. By leveraging a diverse, labeled dataset of diseased leaf images sourced from open repositories, the study established a robust pipeline that included data preprocessing, model training, evaluation, and deployment.

A combination of a custom-built Convolutional Neural Network and transfer learn- ing models such as VGG16, ResNet152V2, and DenseNet were explored. Among these, DenseNet achieved the best performance, reaching an accuracy of 83% after training for 100 epochs. Data augmentation and regularization techniques such as dropout and early stopping were used to improve generalization and mitigate overfitting.

The final model was deployed using a Flask-based API, allowing it to be integrated into a user-friendly web interface. This deployment framework enables real-time predictions by accepting plant leaf images from end users and returning disease classifications with high reliability.

Overall, the project provides a practical and scalable solution that can assist farmers and agricultural experts in identifying plant diseases at an early stage. This not only reduces crop loss but also supports more informed and timely intervention. Future en- hancements may include expanding the dataset to cover more crops and disease types, incorporating multi-language support in the user interface, and integrating weather-based prediction models to further improve accuracy and usability in real-world scenarios.

7. References

1. Singh, G., Guleria, K., & Sharma, S. (2023, October). A Deep Learning-based Fine-tuned Convolutional Neural Network Model for Plant Leaf Disease Detection. In *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)* (pp. 1-6). IEEE.
2. Rashid, M., Ram, B., Batth, R. S., Ahmad, N., Dafallaa, H. M. E. I., & Rehman, M. B. (2019, December). Novel image processing technique for feature detection of wheat crops using python OpenCV. In *2019 International Conference on Compu- tational Intelligence and Knowledge Economy (ICCIKE)* (pp. 559-563). IEEE.
3. Rashid, J., Khan, I., Ali, G., Alturise, F., & Alkhalifah, T. (2023). Real-Time Mul- tiple Guava Leaf Disease Detection from a Single Leaf Using Hybrid Deep Learning Technique. *Computers, Materials & Continua*, 74(1).
4. Mahadevan, K., Punitha, A., & Suresh, J. (2024). A Novel Rice Plant Leaf Diseases Detection Using Deep Spectral Generative Adversarial Neural Network. *Interna- tional Journal of Cognitive Computing in Engineering*.
5. Tandekar, D., & Dongre, S. (2023, June). A Review on Various Plant Disease De- tection Using Image Processing. In *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)* (pp. 552-558). IEEE.
6. Jamal, S., & Judith, J. E. (2023, January). Review on Automated Leaf Disease Pre- diction Systems. In *2023 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)* (pp. 1-4). IEEE.
7. Dorgham, O., Abu-Shareah, G., Alzubi, O., Al Shaqsi, J., Aburass, S., & Al-Betar, M. A. (2024). Grasshopper Optimization Algorithm and Neural Network Classifier for Detection and Classification of Barley Leaf Diseases. *IEEE Open Journal of the Computer Society*.
8. Rashid, R., Aslam, W., & Aziz, R. (2024). An Early and Smart Detection of Corn Plant Leaf Diseases Using IoT and Deep Learning Multi-Models. *IEEE Access*.

9. Gobalakrishnan, N., Pradeep, K., Raman, C. J., Ali, L. J., & Gopinath, M. P. (2020, July). A systematic review on image processing and machine learning techniques for detecting plant diseases. In *2020 international conference on communication and signal processing (ICCSP)* (pp. 0465-0468). IEEE.

10. Barburiceanu, S., Meza, S., Orza, B., Malutan, R., & Terebes, R. (2021). Convo- lutional neural networks for texture feature extraction. Applications to leaf disease classification in precision agriculture. *IEEE Access*, 9, 160085-160103.

11. Dhingra, A., Agarwal, S., Gupta, A. (2019). K-means clustering for segmenta- tion, GLCM for texture features, SVM for classification. *International Journal of Computer Applications*, 178(14), 1-8.

12. Kalaivani, M., Manogaran, G. (2017). K-means clustering for segmentation, Arti- ficial Neural Networks for classification. *International Journal of Pure and Applied Mathematics*, 114(8), 127-133.

13. Vaishnnave, P., Sundararajan, V. (2020). Deep Convolutional Neural Networks (DCNN), Stochastic Gradient Descent with momentum for training. *International Journal of Machine Learning and Computing*, 10(4), 493-500.

14. Xiong, Z., Li, Z., Li, Y. (2020). Modified GrabCut algorithm for segmentation, Deep Learning for classification. *Journal of Computer Vision and Image Processing*, 12(1), 21-32.

15. Sangbamrung, T., Uthayopas, P. (2020). Deep learning for cassava disease classifi- cation. *Journal of Agricultural Informatics*, 11(3), 50-58.

16. Kaur, M., Sharma, A. (2019). Fractional-order Zernike moments for feature extrac- tion, SVM for classification. *International Journal of Computer Vision and Image Processing*, 14(2), 82-92.

17. Vani, M. S., Girinath, S., Hemasree, V., Havardhan, L. H., Sandhya, P. (2023, November). Plant Disease Identification Tracking and Forecasting Using Machine Learning. In *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 1428-1432). IEEE.