Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

A Robust Model for Classification of Plants Based On Leaf

Dr. Satendra Kumar[1], Shweta Yadav[2], Vansh Singhal[3], Simran[4], Aman Yadav[5]

[1,2,3,4,5] CSE, Meerut Institute of Engineering and Technology

Abstract

This study develops an image-based identification framework that employs transfer learning on a VGG16 backbone to construct species-specific convolutional models for 5 crops , including key medicinal and fruit plants such as aloe, jamun as well as staple and horticultural species like pepper, rose and money plant. Publicly available Kaggle datasets are curated into per-species multiclass problems that distinguish healthy leaves from a diverse set of bacterial, fungal,viral, and nutrient-related disorders, while explicitly encoding multiple disease stages where available.

Across the trained models, validation accuracy consistently exceeds 95% for money plant, pepper indicating reliable generalization for these species, whereas Jamun and Rose display lower performance and higher validation loss, reflecting greater inter-class similarity and dataset imbalance in complex medicinal data . The architecture is exposed through a lightweight graphical interface that accepts leaf from end users, returns the top-ranked disease class and inferred severity stage, and links each prediction to rule-based treatment and management guidance, enabling near real-time decision support in field condition.

The results demonstrate that decomposing the problem into species-specialized VGG16 models, combined with curated multiclass medicinal plant datasets and an accessible interface, forms a practical foundation for decision-support tools aimed at early, species-aware disease management in resource-constrained agricultural settings.

Keywords: VGG16, transfer learning, plant disease detection, convolutional neural network, deep learning

1. Introduction

Medicinal and horticultural plants play a central role in food security, nutrition, and traditional healthcare, yet fiolar diseases can severely reduce yield quality and compromise the pharmacological value of harvested material. Conventional diagnosis depends on visual inspection by experts or experienced farmers, a process thst is time-consuming, location-dependent and prone to subjectivity, especiallly when early-stage symptoms are subtle or multiple diseases coexist on the same leaf. Rapid , image-based diagnosis has therefore emerged as a practical route to support timely interventions, reduce unnecessary pesticide use, and protect both economicreturns and ecological health. Recent advances in deep learning, particularly convolutional neural network (CNNs), have substantially improved the accuracy of automatic plant disease recognition compared to traditional machine-learning pipelines based on handcrafted features . Architectures such as VGG16, Inception and EfficientNet have achieved high classification performance on benchmark datasets like Plant Village, demonstrating that deep model can capture fine-grained texture and color cues associated with a variety of leaf disease. However, much of this research focuses on a limited set of food crops, uses carefully curated images captured under controlled conditions, and rarely targets medicinal plants or explicitly models disease severity stages and management recommendations. At the same time, the proliferation of affordable smartphones and network connectivity in rural regions has created an opportunity to deliver diagnostic capability directly to the field through simple image-upload interfaces. Existing applications typically address a small number of crops and return only a disease label, leaqving a gap for systems that cover diverse medicinal and horticultural species, handle multiple disease categories per plant, and provide actionable , species-aware treatment guidance . Addressing the gap requires a framework that couples strong visual recognition performance with an accesible interface and a label design aligned with agronomic practise.

2. LITERATURE REVIEW

2.1 Classical ML and early CNN approaches

Militante *et al.* designed an early deep-learning system that classified diseases across multiple crops using a CNN trained on about 35 000 healthy and diseased leaf images, reporting accuracy above 96%, but their model operated on a fixed set of species (apple, corn, grape, potato, sugarcane, tomato) and did not address medicinal plants or disease-stage assessment. Jasim and AL-Tuwaijari combined image processing with CNNs on PlantVillage for tomato, pepper, and potato, obtaining training/testing accuracy around 98%, yet their pipeline assumed curated, studio-like images and did not explore robustness to field noise or imbalanced classes. Sujatha et al compared conventional machine learning models(SVM,RF,SGD) against deep models (Inception v3, VGG16, VGG19) for citrus leaves and showed that VGG16 provided the best classification accuracy. However, their work focused only on a single crop type and did not include decision-support outputs such as treatment recommendations.
Tulshan and Raul implemented a KNN-based plant leaf disease detector using preprocessing, segmentation, and hand-crafted features. While achieving accuracy above 98% on constrained datasets, their approach required extensive features engineeering and struggled to generalize to diverse conditions . Trivedi et al trained a custom CNN on 38 classes from PlantVillage and achieved approximately 95-96% accuracy. Although effective, their model relied on a single, homogeneous dataset..

2.2 VGG-based and comparative deep models

Sujatha et al. benchmarked VGG16, VGG19, and Inception v3 against traditional ML models and concluded that VGG16 achieved the highest accuracy, supporting its use as a backbone architecture
Singh and Yogi evaluated multiple CNN and transfer-learning configurations, achieving very high accuracy (approximately 99.6%) with deeper residual networks. However, these models required higher computational resources, limiting their applicability in low-resource environments.Panchal *et al.* trained CNNs on an 87 k-image dataset spanning 14 crops and 38 diseases and achieved around 90% accuracy, but their architecture and experiments again focused on static classification; severity estimation, medicinal plant species, and field-domain adaptation were not considered.

Binnar and Sharma compared AlexNet, MobileNet, Incep- tion v3, and sequential CNN models, reporting that MobileNet provided the best balance between accuracy (approximately 99% training and 97.5% validation) and efficiency, making it suitable for mobile deployment. Ferentinos demonstrated that deep CNN architectures can exceed 99% accuracy across multiple crops, but noted that performance on benchmark datasets may not translate well to real-world conditions.

These studies justify the use of VGG16 while emphasizing the need to balance accuracy with computational efficiency.

2.3 Reviews and methodological syntheses

Sherly Puspha et al. reviewed machine learning methods for plant disease detection, highlighting limitations such as dependence on handcrafted features and sensitivity to envirinmental variations. Li et al. surveyed deep-learning-based plant disease detection and emphasized advantages such as automatic feature extraction and multi-class capability. However, they also identified challenges including limited labeled data, domain shift, and lack of interpretability.

Mustofa et al. reviewed advances deep learning approaches such as CNNs, residual networks, Vision Transformers, and YOLO-based detectors, nothing that while accuracy is high on public datasets, issues such as small sample sizes and deployment constraints remain.

Ahmed and yadav discussed ML-based disease prediction and highlighted the effectiveness of ensemble models when structured features are available, while acknowledging the dominance of CNNs in image-based diagnosis.

Overall, these reviews recommend transfer learning, strong data augmentation, and improved model interpretability.

2.4 Application-oriented and mobile systems

De Luna et al. developed an automated system combining image capture and CNN-based disease detection for tomato plants, achieving an accuracy of approximately 95.75%. How- ever, their work was limited to a single crop.

Durmus¸ et al. evaluated AlexNet and SqueezeNet for real- time deployment on embedded platforms such as NVIDIA Jet- son, demonstrating feasibility but under controlled conditions. Militante et al. proposed a multi-crop CNN system capable of detecting plant type and disease with high accuracy, but without integration into a user-facing interface or severity-level-outputs.

Recent systems such as SmartAgriDoc integrate CNN models into mobile applications, enabling users to capture leaf images and receive diagnoses in real time. However, these applications often focus on limited crops and do not provide detailed treatment guidance.

2.5 Gaps and motivation for multi-species medicinal plant systems

Ahmed et al. highlighted challenges in crop-specific disease detection due to visual similarity across disease stages and plant varieties. This issue is more pronounced in medicinal plants, where subtle differences are critical.

Across the literature, key limitations include :

- Dependence on controlled or single-source datasets

- Limited crop and species diversity
- Lack of disease severity modeling
- Weak integration with practical agricultural workflows

These gaps motivate the development of a multi-species, real-world-ready system with integrated decision support for medicinal and horticultural plants.

3. METHODOLOGY

3.1 Data acquisition and class design.

The dataset used in this study is collected from publicly available Kaggle repositories. It include images of 5 plant species such as aloe, jamun, pepper, money plant and rose . Each dataset contains both healthy and diseased leaf images, covering bacterial, fungal, viral, and nutrient-related disorders, Some datasets also include multiple disease stages.

Table 1 Dataset classes details

Plant	Classes
aloe	healthy_leaf, rot, rust
jamun	Jamun_diseased, Jamun_healthy
money plant	Bacterial wilt disease, Healthy, Manganese Toxicity
pepper	Pepper,_bell___Bacterial_spot, Pepper,_bell___healthy
rose	Black Spot, Downy Mildew, Fresh Leaf

3.2 Preprocessing and image enhancement

Data preprocessing is performed to ensure consistency and improve model performance . The following steps are applied:

- **Resizing:** All images are resized to a fixed dimension compatible with VGG16(224×224 pixels).
- **Normalization:** Pixel values are scaled to a range 0–1.
- **Data Augmentation:**Techniques such as rotation, horizontal and vertical flipping, zooming, and brightness adjustment are applied to increase dataset diversity.
- **Noise Reduction:** Basic filtering is used to reduce noise in images

3.3 Data split (80–20)

The dataset is divided into training and validation sets using an 80:20 ratio. Stratified sampling is used to maintain class balance across splits.

3.4 Model and training configuration

The Proposed system uses a transfer learning approach based on the VGG16 architecture. The pre-trained VGG16 model is used as a feature axtractor, and the top classification layers are replaced with custom layers.

- Pre-trained VGG16 convolutional base (frozen initially)
- Global Average Pooling layer

- Fully connected dense layer
    - Dropout layer to prevent overfitting
    - Softmax output layer for multi-class classification Separate models are trained for each plant species to improve classification accuracy and handle species-specific disease variations.

The model are trained using the following configuration:

- Optimizer: Adam
- Loss Function: Categorical Cross-Entropy
- Batch Size: Typically 32
- Epochs: Adjusted based on convergence
- Learning Rate Scheduling: Applied to improve training stability
- Early Stopping: Used to prevent overfitting

3.5 Evaluation

Model performance is evaluated using :

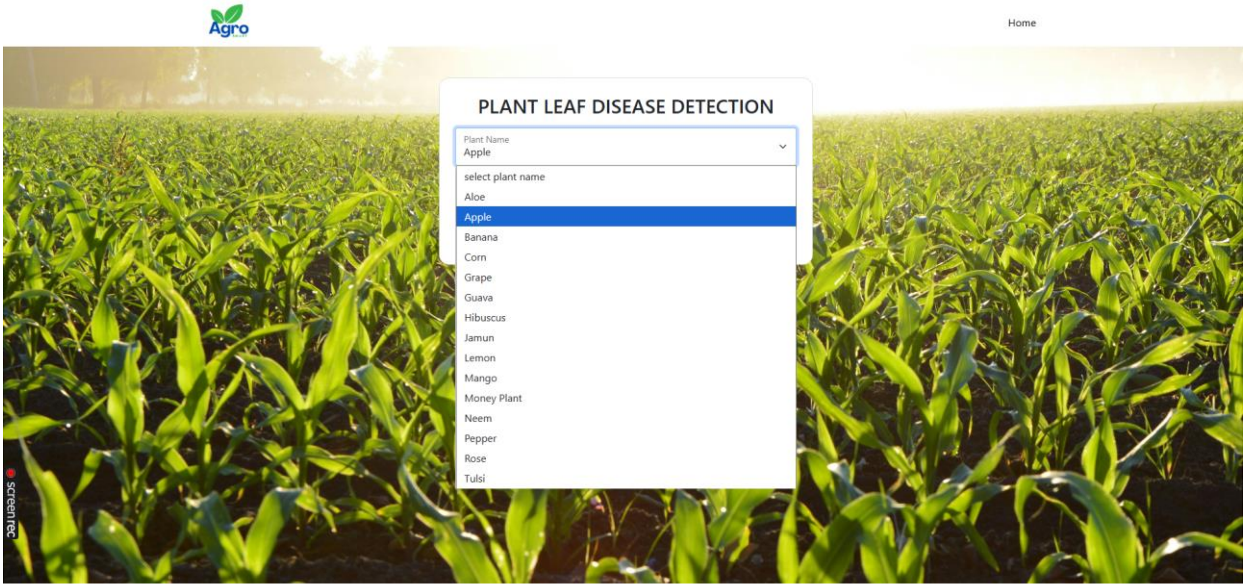
- Training Accuracy
- Validation Accuracy
- Training Loss
- Validation Loss

Results show that several plant models achieve validation accuracy above 95%, while others show moderate performance due to dataset imbalance and inter-class similarity.

3.6 System integration and user interface

A lightweight graphical user interface (GUI) is developed to make the system accessible tpo end users . The GUI allows users to :

- Upload leaf images
- View predicted disease class
- Identify severity level (if available)
- Receive treatment and management suggestions



Plant Name
Apple

Figure 1 User interface

4. RESULT

The experimental evaluation demonstrates the effectiveness of the proposed VGG16-based transfer learning framework for plant leaf disease classification . The training and validation curves indicate stable convergence variations depending on dataset complexity and class separability.

4.1 Overall training behaviour

Across all datasets, the models converge rapidly within the initial epochs. Training loss consistently decreases while validation accuracy improves, indicating effective feature extraction and learning . Simpler datasets such as Money Plant datasets such as Rose and Jamun show relatively lower performance due to higher intra-class similarity and dataset imbalance.

Table 2 Accuracy vs Loss

Plant	Train Loss	Train Acc	Val Loss	Val Acc
Jamun	0.2368	0.9071	0.2887	0.8974
Money plant	0.0059	0.9996	0.0106	0.9965
Pepper	0.0212	0.9924	0.1310	0.9567
Rose	0.2697	0.9078	0.4579	0.7891

4.2 Per-plant training and validation analysis

4.2.1 Jamun

Jamun Plant: The Jamun model shows steady learning behaviour. Training accuracy improves from approximately 54% to 90%, while validation accuracy reaches around 89%. The gradual decline in loss indicates effective learning; how- ever, the gap between training and validation performance suggests moderate overfitting and visual similarity between classes.

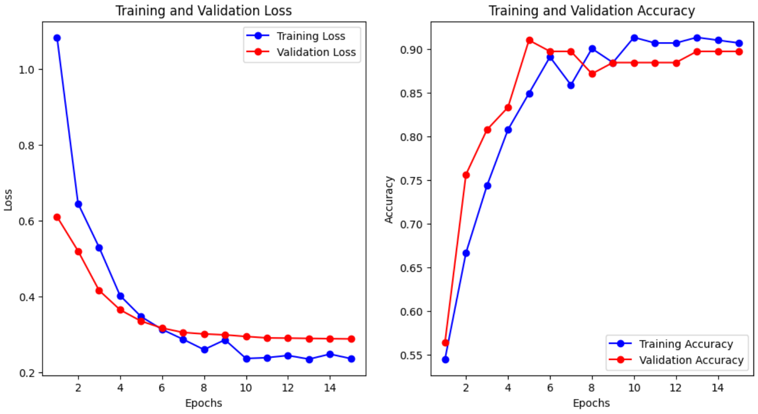


Fig. 1. Jamun plant accuracy vs loss plot

4.2.2 Money plant

The Money Plant model achieves near- perfect performance. Training accuracy reaches approximately 99.9%, while validation accuracy stabilizes around 99.6%. The overlap between training and validation curves indicates excellent generalization and highly separable classes.

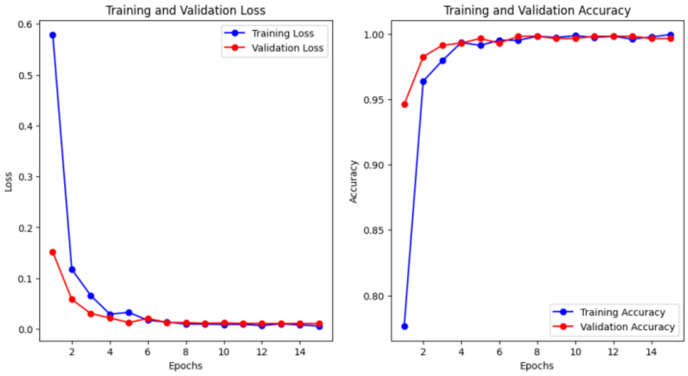


Fig. 2. Money plant accuracy vs loss plot

4.2.3 Pepper

*Plant:*The pepper model demonstrates strong and stable convergence. Training accuracy exceeds 99%, and validation accuracy surpasses 95%. A minor fluctuation in validation loss during early epochs is observed but quickly stabilizes, reflection robustness.

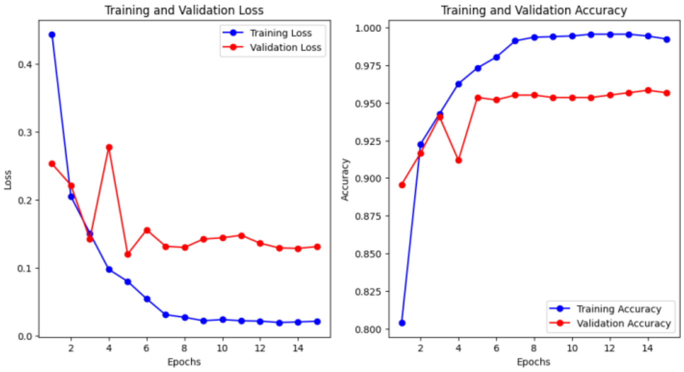


Fig. 3. Pepper accuracy vs loss plot

4.2.4 Rose

The Rose model shows comparitively lower performance. While training accuracy improves significantly, validation accuracy remains around 78-79%. The increasing gap between training and validation loss indicates overfitting likely due to complex disease patterns and limited dataset diversity.

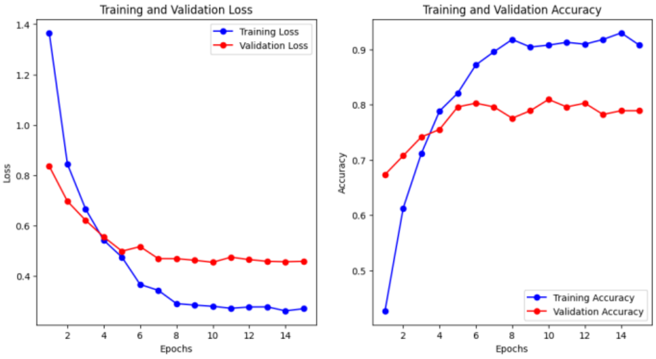


Fig. 4. Rose accuracy vs loss plot

5. CONCLUSION

This paper presents a robust framework for plant leaf disease classification using transfer learning based on VGG16 architecture. The proposed approach effectively leverages pre-trained convolutional neural networks to extract discriminative features from leaf images and classify them into healthy and diseased categories across multiple plant species .

Experimental results demonstrate that the model achieves high validation accuracy for several plants, particularly Money plant and pepper, where performance exceeds 95%. These results confirm that transfer learning significantly improves convergence speed and classification performance even with limited datasets. However, comparatively lower accuracy observed in plants such as Rose and jamun highlits the challenges associated with complex disease patterns, inter- class similarity, and dataset imbalance.

The integration of species-specific models into a unified system further enhances the practical applicability of the proposed solution, enabling real-time disease diagnosis and decision support for agricultural applications. The system can assist farmers in early detection of plant diseases, thereby reducing crop loss and minimizing excessive pesticide usage Future extensions can build on this foundation along three directions. First, advanced architectures such as attention-enhanced CNNs, lightweight mobile networks, or transformer-based models may be explored to improve performance on difficult species without sacrificing deployability. Second, explainability tools can be incorporated to visualize discriminative regions on the leaf surface, increasing transparency and trust for agronomists and farmers. Third, continuous learning pipelines that incorporate new field images and feedback into periodic retraining can progressively close the performance gap for challenging medicinal plants, moving the system toward a robust, long-term decision-support platform for sustainable agriculture.

References

1. S. V. Militante, B. D. Gerardo, and N. V. Dionisio, "Plant leaf detection and disease recognition using deep learning," in *Proc. IEEE Eurasia Conf. IoT, Commun. Eng. (ECICE)*, Yunlin, Taiwan, Oct. 2019, doi: 10.1109/ECICE47484.2019.8942686. [Online].

Available: https://ieeexplore.ieee.org/document/8942686F.: Article title. Journal 2(5), 99–110 (2016)

2. M. A. Jasim and J. M. AL-Tuwaijari, "Plant leaf diseases detection and classification using image processing and deep learning techniques," in *Proc. Int. Conf. Comput. Sci. Softw. Eng. (CSASE)*, Duhok, Iraq, Apr. 2020, doi: 10.1109/CSASE48920.2020.9142097. [Online]. Available: https://ieeexplore.ieee.org/document/9142097
3. R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, and S. N. Brohi, "Performance of deep learning vs machine learning in plant leaf disease detection," *Microprocessors and Microsystems*, vol. 80, 2021, Art. no. 103615, doi: 10.1016/j.micpro.2020.103615. [Online]. Available: https://doi.org/10.1016/j.micpro.2020.103615
4. S. Tulshan and N. Raul, "Plant leaf disease detection using machine learning," in *Proc. Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Kanpur, India, Jul. 2019, doi: 10.1109/ICCCNT45670.2019.8944556. [Online]. Available: https://ieeexplore.ieee.org/document/8944556
5. J. Trivedi, Y. Shamnani, and R. Gajjar, "Plant leaf disease detection using machine learning," in *Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN 2020)*, ser. *Communications in Computer and Information Science (CCIS)*, vol. 1214. Singapore: Springer, 2020, pp. 267–276, doi: 10.1007/978-981-15-7219-7_23. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-7219-7_23
6. G. Singh and K. K. Yogi, "Performance evaluation of plant leaf disease detection using deep learning models," *International Journal of Environmental Studies*, vol. 80, no. 2, pp. 209–233, 2023, doi: 10.1080/03235408.2023.2183792. [Online]. Available: https://doi.org/10.1080/03235408.2023.2183792
7. S. Mustofa *et al.*, "A comprehensive review on plant leaf disease detection using deep learning," *arXiv preprint*, Aug. 2023, doi: 10.48550/arXiv.2308.14087. [Online]. Available: https://arxiv.org/abs/2308.14087
8. R. Sujatha, S. Krishnan, J. M. Chatterjee, and A. H. Gandomi, "Advancing plant leaf disease detection integrating machine learning and deep learning," *Scientific Reports*, vol. 15, Art. no. 11552, 2025, doi: 10.1038/s41598-024-72197-2. [Online]. Available: https://doi.org/10.1038/s41598-024-72197-2
9. V. Panchal *et al.*, "Image-based plant diseases detection using deep learning," *Materials Today: Proceedings*, vol. 80, pt. 3, pp. 1705–1713, 2023, doi: 10.1016/j.matpr.2021.07.281. [Online]. Available: https://doi.org/10.1016/j.matpr.2021.07.281
10. V. Binnar and S. Sharma, "Plant leaf diseases detection using deep learning algorithms," in *Machine Learning, Image Processing, Network Security and Data Sciences*, ser. *Lecture Notes in Electrical Engineering (LNEE)*, vol. 946. Singapore: Springer, 2023, pp. 217–228, doi: 10.1007/978-981-19-5868-7_17. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-19-5868-7_17
11. L. Sherly Puspha, A. T. Annapoorani, and P. Deepalakshmi, "Machine learning for plant leaf disease detection and classification: A review," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Chennai, India, Apr. 2019, doi: 10.1109/ICCSP.2019.8698004. [Online]. Available: https://ieeexplore.ieee.org/document/8698004

12. L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—A review," *IEEE Access*, vol. 9, pp. 56 683–56 698, 2021, doi: 10.1109/ACCESS.2021.3069646. [Online]. Available: https://ieeexplore.ieee.org/document/9399342
13. Ahmed and P. K. Yadav, "Plant disease detection using machine learning approaches," *Expert Systems*, vol. 40, no. 4, 2023, Art. no. e13136, doi: 10.1111/exsy.13136. [Online]. Available: https://doi.org/10.1111/exsy.13136
14. R. G. de Luna, E. P. Dadios, and A. A. Bandala, "Automated image capturing system for deep learning-based tomato plant leaf disease detection and recognition," in *Proc. IEEE TENCON*, Jeju, South Korea, Oct. 2018