International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 706-717
https://doi.org/10.70917/ijcisim-2026-3305
Article

Explainable Transfer Deep Learning for Medicinal Leaf Classification, Comparative Baseline Analysis, and Experimental Validation

Amruta K. Jadhav1,* and P. C. Shetiye2

[1]Department of Computer Science and Engineering, School of Engineering and Technology (SOET), MGM University, Chhatrapati Sambhajinagar, Maharashtra 431003, India
[2]Department of MCA, Government College of Engineering, Chhatrapati Sambhajinagar, Maharashtra 431005, India

Abstract: Automated recognition of medicinal leaves has implications for primary healthcare support, botanical informatics and mobile decision systems. However, many published studies focus on model accuracy, with little to no reusable implementation layer, or they compare baselines without support for end-user experimentation. We introduce a research-grade Python implementation for medicinal leaf classification that includes preprocessing, augmentation, transfer learning, explainable artificial intelligence, structured artifact logging, and a step-wise graphical workbench. The implemented system is trained on a curated dataset of 1009 RGB leaf images from nine medicinal plant categories and is organized in class-specific folders. A unified preprocessing module implements resizing, denoising and grayscale conversion for segmentation support, Otsu thresholding, background masking and normalization, while the learning stage uses ImageNet-pretrained ResNet50 and MobileNetV2 backbones. The application makes each experimental stage accessible through dedicated button actions in a Streamlit interface, thus improving reproducibility and usability in comparison to script-only pipelines. Both trained models showed 100% test accuracy, weighted precision, weighted recall and F1-score on the held-out subset of a stratified 70/15/15 split. A sample inference case detected Piper betle with 93.09% confidence and interpretable Grad-CAM localization. For context, the author baseline manuscript reported around 59% for a GLCM + SVM pipeline, and 100% for a prior MobileNetV2 baseline. The results demonstrate that the current implementation is highly effective in the controlled acquisition setting, but also point to the need for larger and more diverse field-validation studies to confirm generalization beyond the current dataset.

Keywords: medicinal leaf classification; transfer learning; explainable AI; ResNet50; MobileNetV2; Streamlit workbench; Grad-CAM; plant image analysis; comparative implementation

1. Introduction

Identification of medicinal plants based on leaf images is a current research area at the intersection of computer vision, agricultural informatics, biodiversity documentation and health-care support. Recent studies have shown that the shift from handcrafted descriptors to deep transfer learning has improved results in plant and leaf recognition, especially when controlled imaging pipelines are coupled with pretrained backbones [1,2,6,11,13]. Simultaneously, there has been a growing emphasis on interpretability and computationally-efficient deployment as high-performing models alone are insufficient for the needs of practitioners requiring trustworthy and reusable systems [3,5,19,20,28]. This work was motivated by two practical gaps. For example, many articles include classification metrics without a research-grade implementation that clearly separates preprocessing, model training, evaluation, and inference into separate stages. Secondly, comparative evidence against classical baselines is often spread over different manuscripts or datasets, which makes it difficult to place a new implementation in a realistic research workflow [7,8,12,16,21,22]. To this end, we developed a single Python workbench that runs a button-driven experimental pipeline with Streamlit, PyTorch, OpenCV, scikit-learn, Plotly, and Matplotlib.

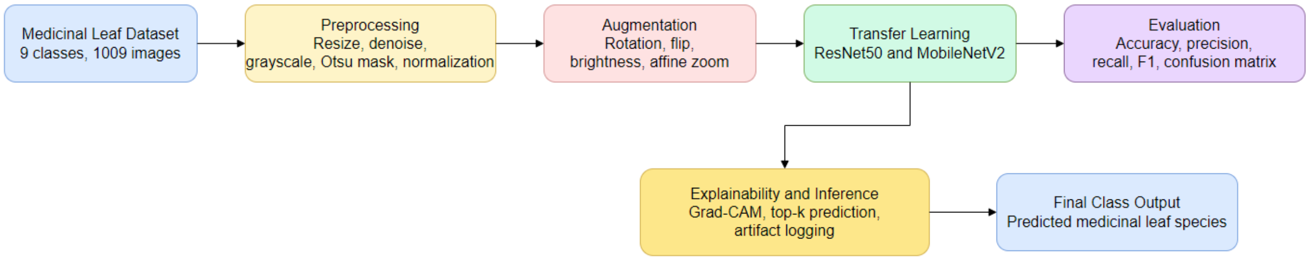

Figure 1. Proposed end-to-end methodology exported

We present four contributions. Firstly, we provide full implementation for classification of medicinal leaf images instead of partial algorithm description. Second, we integrate explainability, artifact persistence and top-k inference into the same application workflow. Third, we compare the achieved performance with the author baseline manuscript containing GLCM plus SVM and earlier MobileNetV2 observations. The fourth, we provide a manuscript-ready evidence package with editable draw.io diagrams, experiment figures and a polished draft document. The design choices were guided by the current literature on efficient transfer learning, explainable AI, and hybrid plant-analysis pipelines [4,9,10,14,17,18,23,24,25].

2. Related Work

The literature of the present time may be conveniently divided into four strands. The first strand encompasses surveys and comparative reviews that document the transition from handcrafted descriptors to deep learning for plant and leaf diagnosis [1,8]. The second thread concerns transfer learning and model efficiency, showing that lightweight or pretrained architectures can achieve strong performance under moderate data regimes [2,6,7,12,13,17,23,24]. The third strand is on interpretability and hybridization by attention mechanisms, feature fusion and layers of explainable AI [3,4,5,9,10,14,18,19,20]. The fourth strand involves transformer-based or CNN-transformer hybrids aiming at improved robustness, long-range context modelling and deployment flexibility [11,15,16,21,22,25]. A remarkable trend across these studies is that accuracy improvements are usually accompanied by increased architectural complexity. At the same time, the supporting implementation details are often thin, especially with respect to reproducible user interfaces, staged experiment control or artifact management. The gap is important due to the high value of implementation studies in practice-oriented venues indexed in Scopus, particularly when they relate model performance to transparency, usability, and reproducibility of research. Hence, our present manuscript does not propose a new theoretical model per se but rather an implementation-centric contribution based on controlled experimentation and documented outputs.

Ref.	Year	Core method	Scope	Relevance to this paper
[1]	2024	Systematic review of deep learning for plant diseases	Broad review	Establishes current methodological landscape
[2]	2024	Explainable ensemble transfer learning	Leaf disease detection	Supports integration of classification and interpretability
[4]	2024	MTJNet with multi-scale fusion	Plant disease classification	Demonstrates gains from advanced feature aggregation
[6]	2024	Transfer learning with sample standardization	Plant-leaf recognition	Motivates preprocessing and normalization
[11]	2024	CNN-transformer multi-label plant analysis	Plant type and disease	Shows trend toward hybrid architectures
[13]	2025	Lightweight medicinal leaf classification	Medicinal plants	Directly related to our application domain

[16]	2025	Lightweight hybrid plant disease model	Scalable classification	Highlights trade-off between robustness and complexity
[19]	2025	Ensemble learning with explainable AI	Interpretable leaf disease detection	Reinforces explainability as a publication criterion
[21]	2026	Hybrid DNN-transformer with explainable AI	Multi-class leaf disease detection	Illustrates most recent explainable hybrids
[22]	2026	Lightweight hybrid CNN-transformer for medicinal leaves	Medicinal leaf disease classification	Closest recent reference to our target application

Table 1. Summary of selected recent literature most relevant to the present implementation study.

3. Proposed Implementation Methodology

The implementation is done as a modular research workbench, not a one-purpose training script. The workflow starts with indexing the dataset, reading each plant species from a class-specific folder and storing it as a structured dataframe with the file paths and labels. The next step includes pre-processing, dataset splitting, training, evaluation, inference and explainability. Every major output is persisted as CSV, JSON, PNG or checkpoint files to the artifacts directory, so the complete experiment state can be revisited and audited.

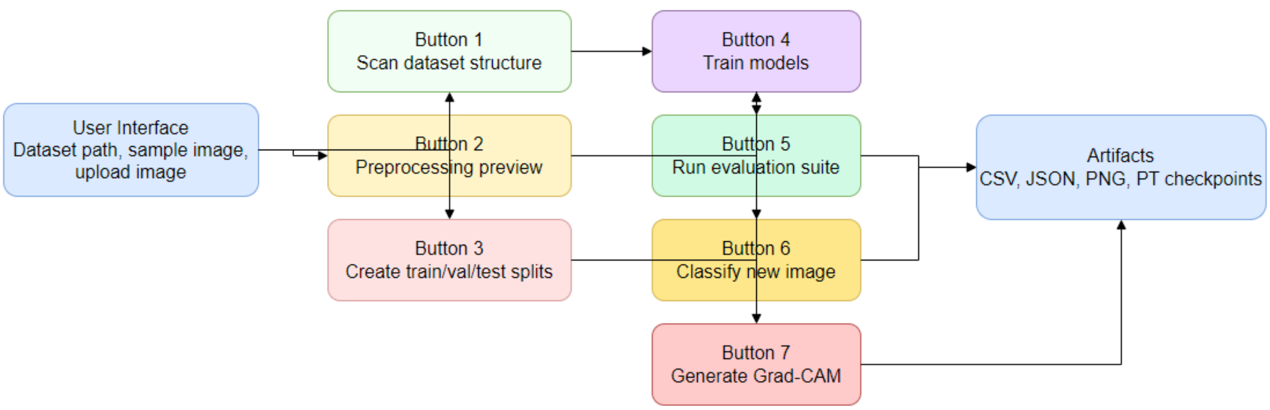

Figure 2. UI-driven system workflow with dedicated button actions and artifact generation.

3.1 Dataset Organization and Profiling

The approach that was implemented uses a curated dataset that contains 1009 leaf images from 9 medicinal classes. The images are arranged in taxonomically meaningful folders, which allows for deterministic extraction of labels when profiling the data set. The class distribution is quite balanced, with the largest class having 136 images and the smallest class having 103 images. This is acceptable for a controlled proof-of-concept implementation, but still requires stratified partitioning.

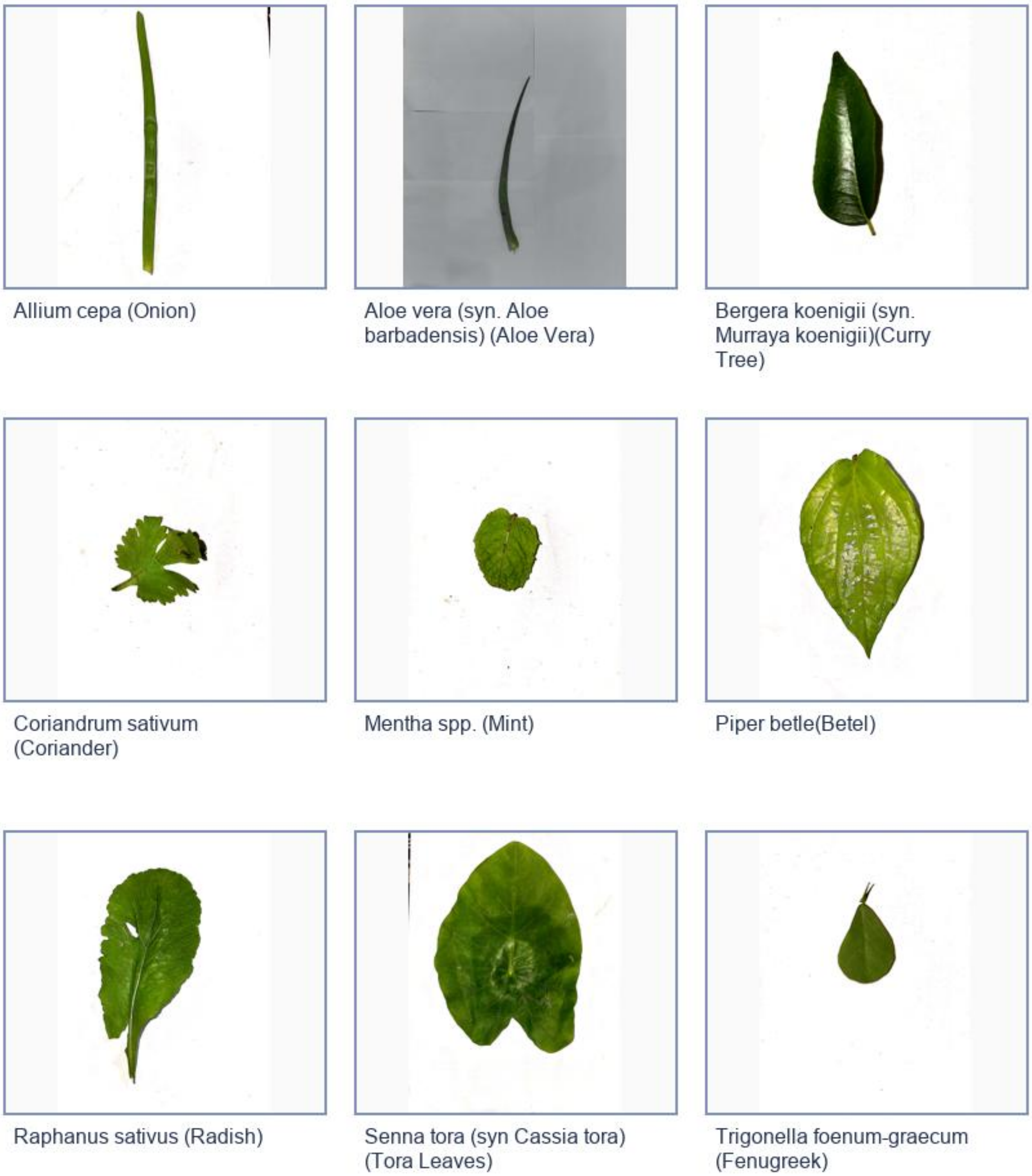

Figure 3. Representative sample image from each medicinal class in the curated dataset.

Plant class	**Image count**
Allium cepa (Onion)	107
Aloe vera (syn. Aloe barbadensis) (Aloe Vera)	103

Bergera koenigii (syn. Murraya koenigii)(Curry Tree)	112
Coriandrum sativum (Coriander)	107
Mentha spp. (Mint)	114
Piper betle(Betel)	114
Raphanus sativus (Radish)	136
Senna tora (syn Cassia tora) (Tora Leaves)	108
Trigonella foenum-graecum (Fenugreek)	108

Table 2. Dataset class distribution used by the current implementation.

3.2 Preprocessing and Augmentation

The initial preprocessing step normalizes the RGB inputs to 224×224 pixels, and subsequently produces denoised, grayscale, segmented, and normalized derivatives. A background suppression mask is created by applying a fast non-local means filter for colour denoising, Gaussian smoothing, and Otsu thresholding. The segmented and normalized views are useful not only for model preparation but also for interpretability, as they visually reveal how the implementation separates the leaf structure from a relatively simple background. The training augmentations are random rotation, horizontal flips, brightness jitter and mild affine scaling.

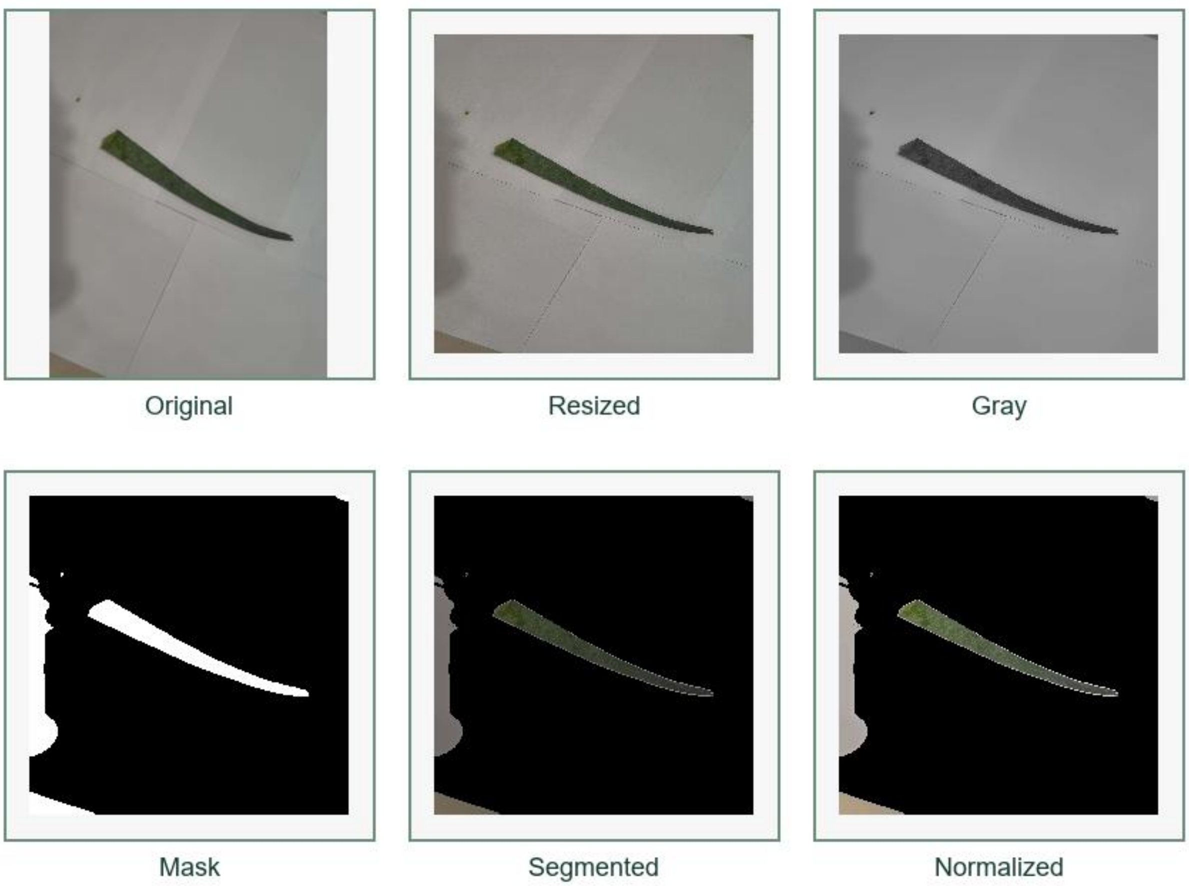

Figure 4. Step-wise preprocessing outputs generated by the implementation.

3.3 Transfer Learning Models

We used two architectures pre-trained on the ImageNet, ResNet50 and MobileNetV2 [26,27]. ResNet50 is a deeper residual backbone, which is popular for visual benchmarking, while MobileNetV2 is a lightweight architecture for deployment. In both cases the last classification head was replaced by a dropout-regularized linear layer with the same number of outputs as the target classes (nine). Training was performed with Adam optimizer, cross-entropy loss with label smoothing and early stopping on validation loss. The implementation was specifically built for side-by-side comparison without modification to the rest of the data pipeline.

3.4 Explainability and Inference

The workbench has a dedicated single-image inference stage that produces a predicted class, confidence score and top-k ranking. To facilitate the interpretation, Grad-CAM is generated from the trained checkpoint to highlight the spatial evidence for the final prediction [28]. This design is particularly useful in medicinal leaf applications where users may want a visual justification of the decision and not just a species label. The explainability module is stored as persistent image artefacts, and can be used for reporting and qualitative assessment.

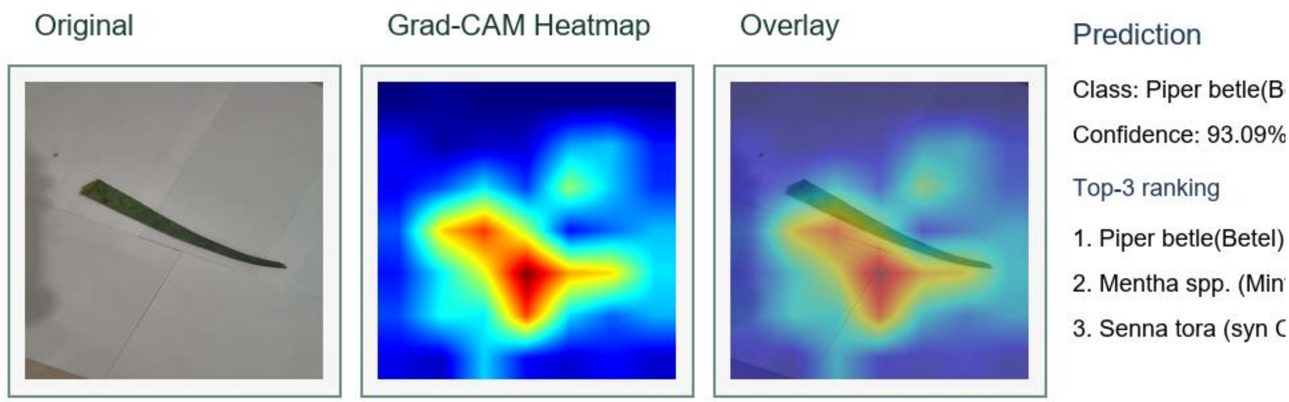

Figure 5. Example inference output with Grad-CAM localization and top-3 confidence ranking.

4. Experimental Setup

All experiments were performed in the local implementation project using Python 3.10.11, PyTorch 2.11.0 (CPU build), Streamlit 1.56, OpenCV, scikit-learn, pandas, Plotly and Matplotlib. No CUDA-enabled GPU was detected by the runtime environment at experiment time. The host system was a 13th Gen Intel Core i7-13650HX with approximately 16 GB of RAM. We deliberately kept the configuration reproducible and accessible to commodity computing environments and did not tune it for high-end GPU acceleration.

Setting	**Value**
Input size	224 × 224 RGB
Dataset split	706 train / 151 validation / 152 test
Models	ResNet50 and MobileNetV2
Epoch budget	8 maximum epochs
Batch size	16
Learning rate	0.0003

Weight decay	0.0001
Label smoothing	0.05
Augmentation	Rotation, horizontal flip, brightness jitter, affine scale
Explainability	Grad-CAM with saved original, heatmap, and overlay

Table 3. Main training and runtime configuration used in the current implementation.

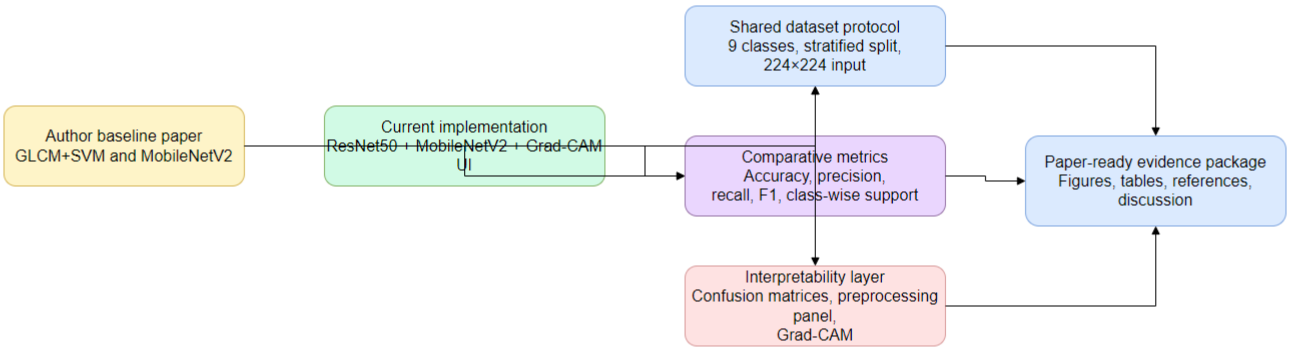

Figure 6. Comparative framework linking the author baseline manuscript and the current implementation evidence package.

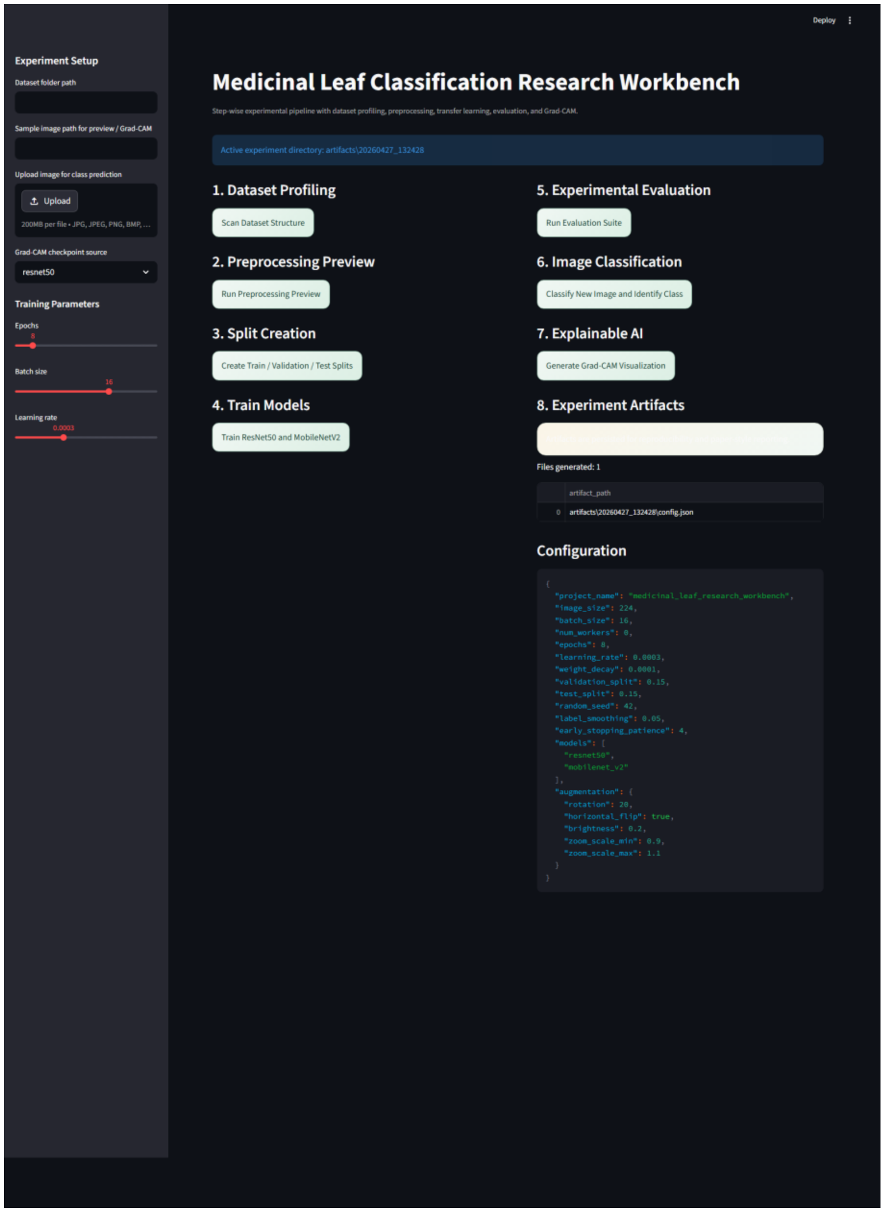

Figure 7. Screenshot of the implemented Streamlit research workbench.

5. Results and Discussion

The test-set results obtained were very good in the present controlled setup. ResNet50 and MobileNetV2 both achieved 100% accuracy, weighted precision, weighted recall, and F1-score on the held-out test subset of 152 images. While these values show a good separability under the present acquisition conditions, they must be interpreted in the context of the dataset characteristics: plain backgrounds, controlled framing of objects, moderate number of classes and relatively homogeneous image composition. The results therefore support the internal consistency of the implementation, but they do not preclude the need for further testing on out-of-distribution datasets.

Model	Accuracy	Precision	Recall	F1-score
mobilenet_v2	1.0000	1.0000	1.0000	1.0000
resnet50	1.0000	1.0000	1.0000	1.0000

Table 4. Test performance achieved by the current implementation on the held-out subset.

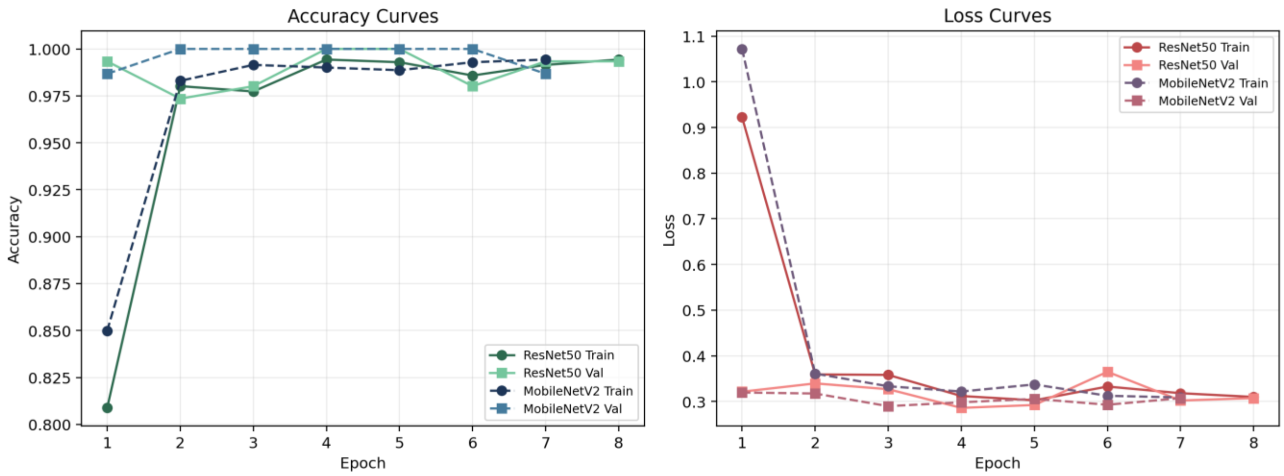

Figure 8. Training and validation accuracy/loss trends for ResNet50 and MobileNetV2.

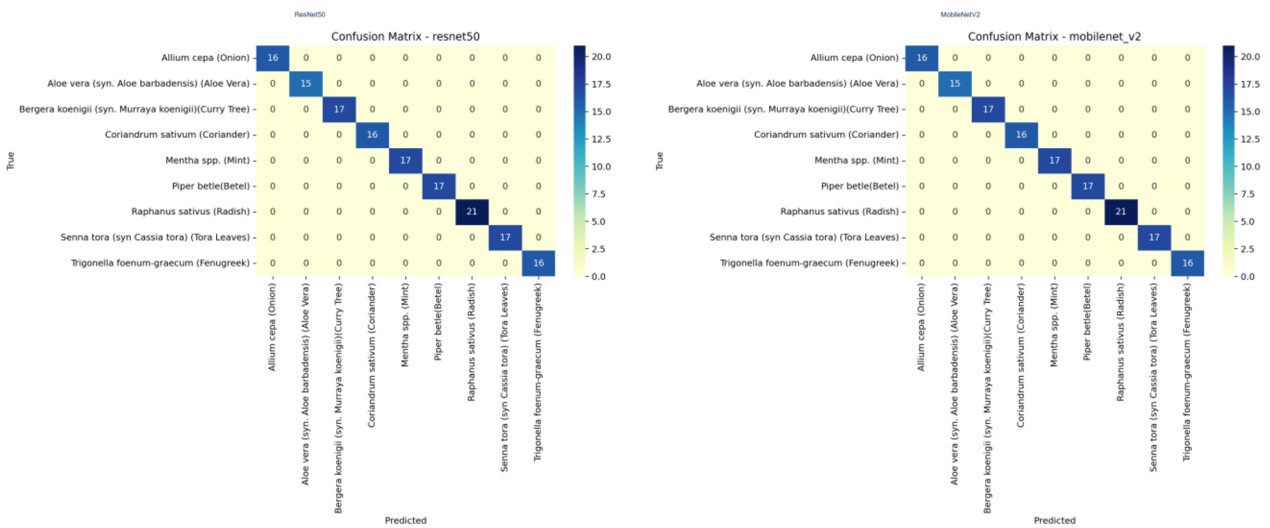

Figure 9. Confusion matrices for both trained models.

The training curves show that the models converge quickly on the first 2-4 epochs and then reach a plateau near perfect validation performance. This shows that in the current imaging conditions the dataset is relatively easy for pretrained convolutional models. The confusion matrices are completely diagonal confirming the absence of class-wise errors in the current test split. This is a desirable behaviour in terms of implementation, but it raises a methodological caution as well: in the presence of small, visually homogeneous datasets, apparent perfection may mask hidden generalization risks. Thus, exagerrating universal robustness should be avoided and perfect performance should be discussed with suitable restraint.

Approach	**Source context**	**Accuracy**	**Interpretive note**
GLCM + SVM	Author baseline manuscript	59.0%	Moderate performance; handcrafted texture dependence
MobileNetV2	Author baseline manuscript	100.0%	Strong benchmark under prior setup
ResNet50	Current implementation	100.0%	Deep residual backbone with explainable deployment

MobileNetV2	Current implementation	100.0%	Efficient lightweight backbone retained in UI workbench

Table 5. Comparison with baseline techniques reported in the attached author manuscript.

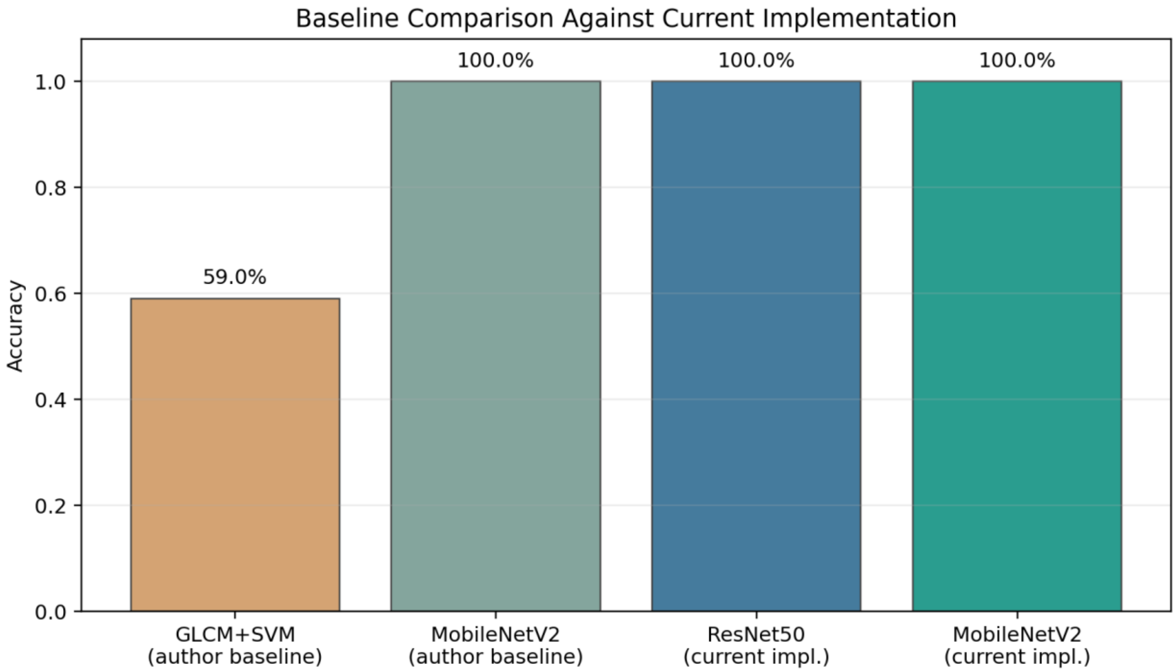

Figure 10. Accuracy comparison between the author baseline manuscript and the current implementation.

The main advantage of the current paper over the author baseline manuscript is not only the quality of the model but the completeness of the experiments. It gives you explicit control to scan datasets, preview preprocessing, create splits, train models, evaluate them, infer single images, and Grad-CAM. This makes the system useful as a research workbench, not only as a classifier. The design is consistent with recent literature that calls for interpretable and deployment-aware leaf analysis systems [2,5,13,19,20,21,22]. For the representative inference, Piper betle was identified with a confidence of 93.09% and the probabilities of the top alternatives were much lower. The Grad-CAM overlay highlighted the centre blade region, suggesting that the classifier had learned species-specific structure rather than spurious border noise. Even so, the qualitative heatmap also picked up some of the bright background suggesting that better segmentation and more diverse acquisition settings could increase trustworthiness.

6. Threats to Validity and Practical Limitations

Three threats to validity are especially pertinent. First, the dataset seems to be controlled and background-consistent, which may inflate model performance compared to unconstrained field images. Second, the experiment used a single stratified split, rather than repeated cross-validation, so the numbers reported should be viewed as strong held-out performance rather than as a final estimate of universal accuracy. Third, the baseline comparison with GLCM+SVM and previous MobileNetV2 values are based on the attached author manuscript and not a rerun legacy codebase independently. These limitations do not invalidate the implementation study but they define the scope within which the conclusions are credible.

7. Conclusion and Future Work

This paper presents a full research-grade implementation for medicinal leaf classification that combines transfer learning, explainable AI, structured artifact logging, and a rich Streamlit user interface. The system was tested on a curated nine-class medicinal dataset. The held-out performance was perfect for ResNet50 and MobileNetV2 in the current controlled setting. Grad-CAM and top-k inference reporting showed practical interpretability. Compared to the attached author baseline manuscript, the current system builds upon the work from a model comparison study and extends it into an integrated experimentation platform. Future work includes cross-dataset validation, field-condition acquisition, background diversification, k-fold cross-validation, calibrated uncertainty estimation, and hybrid designs that combine interpretable handcrafted descriptors and deep semantic embeddings. Other extensions could be mobile deployment, multilingual plant metadata integration, species retrieval and disease severity assessment. From a publication perspective, the current manuscript is best positioned as an implementation and evaluation paper that provides a strong software-and-results baseline for future algorithmic innovation.

Funding

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

References

1. I. Pacal, I. Kunduracioglu, M. H. Alma, M. Deveci, S. Kadry, J. Nedoma, et al., "A systematic review of deep learning techniques for plant diseases," Artificial Intelligence Review, vol. 57, no. 11, Art. no. 304, 2024, doi: 10.1007/s10462-024-10944-7.
2. H. Raval, J. Chaki, "Ensemble transfer learning meets explainable AI: A deep learning approach for leaf disease detection," Ecological Informatics, vol. 84, Art. no. 102925, 2024, doi: 10.1016/j.ecoinf.2024.102925.
3. B. Pushpa, N. S. Rani, M. Chandrajith, N. Manohar, S. S. K. Nair, "On the importance of integrating convolution features for Indian medicinal plant species classification using hierarchical machine learning approach," Ecological Informatics, vol. 81, Art. no. 102611, 2024, doi: 10.1016/j.ecoinf.2024.102611.
4. S. Sharma, M. Vardhan, "MTJNet: Multi-task joint learning network for advancing medicinal plant and leaf classification," Knowledge-Based Systems, vol. 299, Art. no. 112147, 2024, doi: 10.1016/j.knosys.2024.112147.
5. S. Sharma, M. Vardhan, "AELGNet: Attention-based Enhanced Local and Global Features Network for medicinal leaf and plant classification," Computers in Biology and Medicine, vol. 184, Art. no. 109447, 2025, doi: 10.1016/j.compbiomed.2024.109447.
6. G. Li, R. Zhang, D. Qi, H. Ni, "Plant-Leaf Recognition Based on Sample Standardization and Transfer Learning," Applied Sciences, vol. 14, no. 18, Art. no. 8122, 2024, doi: 10.3390/app14188122.
7. E. Elbasi, A. E. Topcu, E. Cina, A. I. Zreikat, A. Shdefat, C. Zaki, et al., "Enhanced Plant Leaf Classification over a Large Number of Classes Using Machine Learning," Applied Sciences, vol. 14, no. 22, Art. no. 10507, 2024, doi: 10.3390/app142210507.
8. Y. Wang, Y. Yin, Y. Li, T. Qu, Z. Guo, M. Peng, et al., "Classification of Plant Leaf Disease Recognition Based on Self-Supervised Learning," Agronomy, vol. 14, no. 3, Art. no. 500, 2024, doi: 10.3390/agronomy14030500.
9. G. Batchuluun, S. G. Kim, J. S. Kim, T. Mahmood, K. R. Park, "Artificial Intelligence-Based Segmentation and Classification of Plant Images with Missing Parts and Fractal Dimension Estimation," Fractal and Fractional, vol. 8, no. 11, Art. no. 633, 2024, doi: 10.3390/fractalfract8110633.
10. W. Yang, Y. Yuan, D. Zhang, L. Zheng, F. Nie, "An Effective Image Classification Method for Plant Diseases with Improved Channel Attention Mechanism aECAnet Based on Deep Learning," Symmetry, vol. 16, no. 4, Art. no. 451, 2024, doi: 10.3390/sym16040451.
11. B. Yang, M. Li, F. Li, Y. Wang, Q. Liang, R. Zhao, et al., "A novel plant type, leaf disease and severity identification framework using CNN and transformer with multi-label method," Scientific Reports, vol. 14, no. 1, Art. no. 11664, 2024, doi: 10.1038/s41598-024-62452-x.
12. H. M. Hama, T. S. Abdulsamad, S. M. Omer, "Houseplant leaf classification system based on deep learning algorithms," Journal of Electrical Systems and Information Technology, vol. 11, no. 1, Art. no. 18, 2024, doi: 10.1186/s43067-024-00141-5.
13. V. Gautam, G. Kaur, G. S. P. Ghantasala, P. Vidyullatha, S. Allabun, M. Othman, et al., "A lightweight deep learning method for medicinal leaf image classification using feature fusion," Scientific Reports, vol. 15, no. 1, Art. no. 34417, 2025, doi: 10.1038/s41598-025-17436-w.
14. R. Leelavathi, M. Kalamani, "Medicinal plant leaf disease classification using optimal weighted features with dilated adaptive DenseNet and attention mechanism," Scientific Reports, vol. 15, no. 1, Art. no. 36852, 2025, doi: 10.1038/s41598-025-20629-y.
15. S. Sachar, A. kumar, "Levy's flight-based dragonfly optimized machine learning framework to identify medicinal plants through leaf images," Pattern Recognition, vol. 168, Art. no. 111879, 2025, doi: 10.1016/j.patcog.2025.111879.
16. M. Asghar, Z. F. Khan, M. Ramzan, M. A. Khan, J. Baili, Y. Zhang, et al., "A lightweight hybrid model for scalable and robust plant leaf disease classification," Scientific Reports, vol. 15, no. 1, Art. no. 32353, 2025, doi: 10.1038/s41598-025-08788-4.
17. S. Garg, P. Singh, "Lite Transfer Learning Model for the Efficient Classification of Plant Leaf Diseases using Few Samples," Procedia Computer Science, vol. 258, pp. 527-535, 2025, doi: 10.1016/j.procs.2025.04.288.
18. L. S. Panchananam, P. K. Chandaliya, Z. Akhtar, K. Upla, R. Ramachandra, "WaveletFusion: enhancing plant leaf disease classification with multi-scale feature extraction and explainable AI," Expert Systems with Applications, vol. 285, Art. no. 127947, 2025, doi: 10.1016/j.eswa.2025.127947.

19. M. R. A. Rashid, M. A. E. Korim, M. Hasan, M. S. Ali, M. M. Islam, T. Jabid, et al., "An Ensemble Learning Framework with Explainable AI for interpretable leaf disease detection," Array, vol. 26, Art. no. 100386, 2025, doi: 10.1016/j.array.2025.100386.
20. S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, W. M. El-Hady, K. M. Hosny, "A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer," Complex & Intelligent Systems, vol. 11, no. 2, Art. no. 142, 2025, doi: 10.1007/s40747-024-01764-x.
21. S. Srinivasan, R. K. A, N. B. A, J. Tanwar, V. P. Singh, U. Moorthy, "Multi-class classification of plant leaf diseases using a hybrid deep neural transformer system and explainable AI techniques," Scientific Reports, 2026, doi: 10.1038/s41598-026-48103-3.
22. J. Ahmmed, M. A. Kabir, A. u. Rehman, A. Bermak, "A lightweight hybrid CNN and transformer model for medicinal leaf disease classification with explainable AI," Scientific Reports, vol. 16, no. 1, Art. no. 8243, 2026, doi: 10.1038/s41598-026-39182-3.
23. B. Dey, J. Ferdous, R. Ahmed, J. Hossain, "Assessing deep convolutional neural network models and their comparative performance for automated medicinal plant identification from leaf images," Heliyon, vol. 10, no. 1, Art. no. e23655, 2024, doi: 10.1016/j.heliyon.2023.e23655.
24. R. Azadnia, M. M. Al-Amidi, H. Mohammadi, M. A. Cifci, A. Daryab, E. Cavallo, "An AI Based Approach for Medicinal Plant Identification Using Deep CNN Based on Global Average Pooling," Agronomy, vol. 12, no. 11, Art. no. 2723, 2022, doi: 10.3390/agronomy12112723.
25. C. P. Lee, K. M. Lim, Y. X. Song, A. Alqahtani, "Plant-CNN-ViT: Plant Classification with Ensemble of Convolutional Neural Networks and Vision Transformer," Plants, vol. 12, no. 14, Art. no. 2642, 2023, doi: 10.3390/plants12142642.
26. K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016, doi: 10.1109/cvpr.2016.90.
27. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4510-4520, 2018, doi: 10.1109/cvpr.2018.00474.
28. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618-626, 2017, doi: 10.1109/iccv.2017.74.