International Research Journal of Multidisciplinary Scope (IRJMS), 2026;7(3):185-200 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i03.09761

An Innovative Method for Plant Species Classification Utilizing Convolutional Neural Networks

Karnan A*, Ragupathy R

Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Chidambaram, Tamil Nadu, India. *Corresponding Author's Email: mekarnan@gmail.com

Abstract

Accurate plant species identification is essential for biodiversity monitoring, ecological analysis and automated environmental assessment. However, conventional deep learning models often struggle to distinguish visually similar plant species due to subtle morphological variations and high inter-class similarity. To address this limitation, this study presents a hierarchical deep learning architecture integrating ResNet50 feature extraction, k-means clustering and cluster-specific Convolutional Neural Network (CNN) classifiers for fine-grained plant species classification. ResNet50 is employed to extract discriminative morphological representations, while k-means clustering organizes visually related species into compact feature groups before classification, thereby reducing inter-class confusion and enabling specialized CNN learning within each cluster. The proposed methodology was evaluated using the PlantCLEF 2015 dataset containing 113,205 images representing 1,000 plant species, including trees, herbs and ferns from Western Europe. Experimental evaluation achieved 0.9800 classification accuracy with 0.0615 loss, outperforming conventional architectures including ResNet50, InceptionV3, Xception, EfficientNet and MobileNet. In addition, the proposed architecture maintained efficient computational performance with an average inference latency of 18 ms per image. These findings demonstrate the suitability of the architecture for automated biodiversity assessment, ecological monitoring and large-scale botanical image analysis. Future work will focus on lightweight attention-enhanced architectures and cross-dataset generalization for deployment in resource-constrained environments.

Keywords: Convolutional Neural Networks, Deep Learning Methodology, K-means Clustering, Plant Species Classification, ResNet50.

Introduction

Reliable identification of plant species is fundamental for ecological monitoring, conservation planning and automated biodiversity assessment. While accurate classification is fundamental to establishing effective conservation strategies, conventional identification techniques frequently falter in complex natural habitats due to limited robustness (1). Historically, early methodologies relied on digital image refinement, manual feature derivation and basic statistical learning models. Although these efforts laid the groundwork for digital botany, they often lacked the flexibility needed to handle complex datasets. For instance, Support Vector Machines (SVMs) that integrate structural descriptors with textural data are often hampered by non-standardized feature extraction and high sensitivity to dataset variations (2). Similarly, traditional neural networks that rely on morphological traits often struggle to capture the intricate structural diversity of leaf patterns (3, 4). The evolution of neural architecture has transformed botanical identification, shifting the paradigm toward end-to-end deep learning frameworks. Residual learning models and deep convolutional architectures have demonstrated substantial improvements in classification performance, with architectures such as Visual Geometry Group 16 (VGG16) reporting accuracy rates of up to [95%] on benchmark datasets (5). Furthermore, hybrid systems that merge spatial feature learning with Bidirectional Long Short-Term Memory (Bi-LSTM) units have optimized feature fusion, resulting in greater stability (6). Modern innovations, such as the Visible-Infrared Convolutional Neural Network (VI-CNN) architecture, which synthesizes hyperspectral Light Detection and Ranging (LiDAR) signals, have demonstrated that incorporating spectral data can enhance classification reliability (7). Additionally, systems like D-Leaf underscore the advantages of automated feature learning over manual

(Received 19th December 2025; Accepted 02nd June 2026; Published 02nd July 2026)

engineering (8). As network topologies become more sophisticated, classification benchmarks continue to improve (9-11).

Despite these advancements, a significant hurdle remains: identifying species from multi-image datasets. Consequently, this study reformulates the classification challenge as an identification task using multi-image observation queries-each query comprising multiple photographs depicting different views or aspects of a single plant specimen (e.g., leaf venation, flower morphology, stem texture). This mirrors real-world scenarios in biodiversity monitoring, where field researchers capture multiple images of the same plant under varying lighting conditions and in precision agriculture, where farmers photograph crops from different angles to identify diseases or pest damage. To overcome existing constraints, introduces a multi-stage deep learning framework for fine-grained plant species classification. The pipeline begins with Residual Network 50 (ResNet50), which is used to derive discriminative feature representations from the input imagery (12). These features are subsequently clustered using k-means-an organ-type-driven selection process that groups images based on visual similarity, which often corresponds to different plant organs (leaves, flowers, stems) or viewing perspectives. The final taxonomic assignment is performed by a cascaded series of Convolutional Neural Networks (CNNs), which excel at interpreting sophisticated visual characteristics (13, 14). This hierarchical structure allows specialized networks to focus on distinct visual attributes while multi-resolution analysis captures fine-grained morphological details (15, 16).

Methodology

This segment outlines the conceptual framework and the systematic execution path of the introduced plant identification system. The architecture is organized into three distinct operational modules: feature extraction using ResNet50, feature grouping through k-means clustering and final classification performed by Convolutional Neural Networks (CNNs).

System Architecture

The proposed framework integrates three main processing stages: feature extraction, feature clustering and species classification. Before feature extraction, data augmentation is applied to increase sample diversity across all plant categories. The augmented images are then processed through a standardised preprocessing pipeline comprising quality checking, pixel-intensity normalisation to the [0, 1] range and spatial resizing to 224 × 224 pixels, as illustrated in Figure 1.

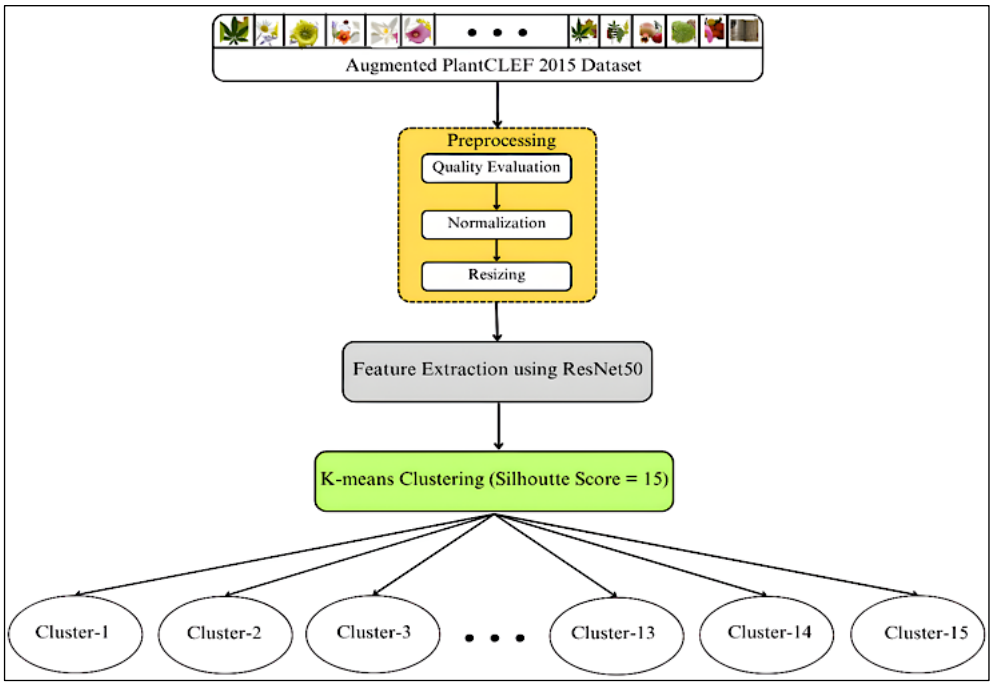

Figure 1: Feature Extraction and Clustering

The dataset was partitioned into three stratified subsets: 60% for training, 20% for validation and 20% for final testing. Stratification was applied to preserve the original class distribution across all subsets. Five-fold cross-validation was then conducted within the training portion to assess model stability. In each fold, four subsets were used for training and one subset for validation, rotating across all five combinations. The cross-validation procedure was non-nested and the independent test set remained fully isolated during model training, validation and hyperparameter selection. During training, a pre-trained ResNet50 network is used to extract high-dimensional visual descriptors from each image (12). ResNet50 functions as a feature encoder, converting raw pixel data into compact 2,048-dimensional vectors that capture morphological characteristics such as leaf shape, venation and surface texture. These vectors are then passed to the clustering stage, where visually similar samples are grouped before classification. The feature space is partitioned into 15 clusters based on Silhouette Score analysis, as detailed in the clustering subsection.

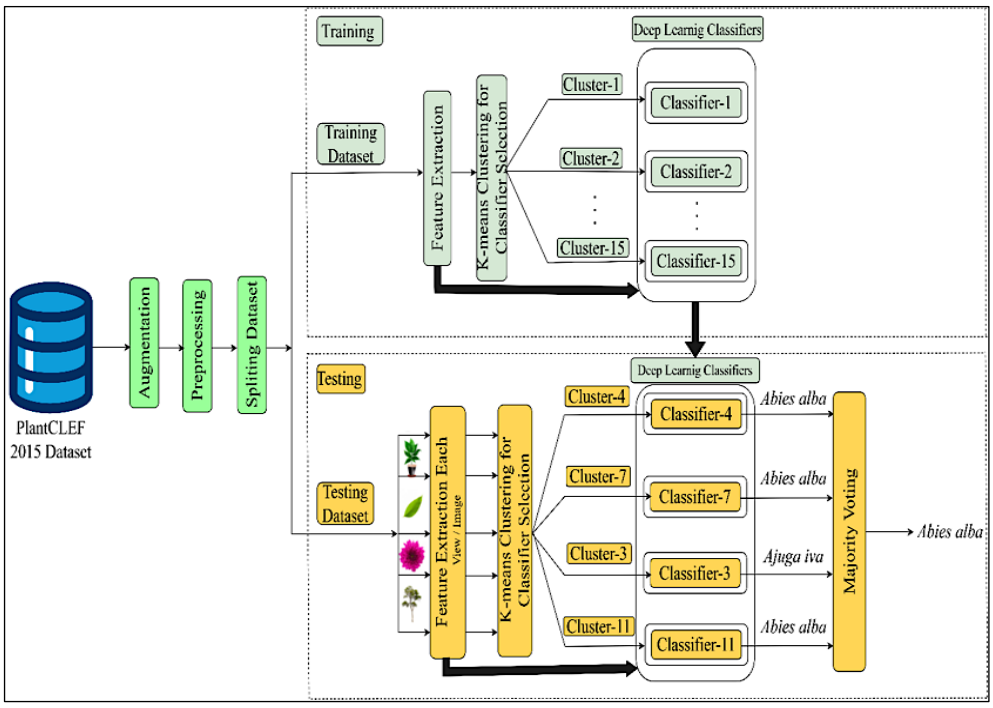

Figure 2: System Architecture for Plant Species Classification

This intermediate grouping reduces inter-class confusion by allowing each downstream CNN to focus on a narrower and more homogeneous subset of plant species. Rather than training a single global classifier across all 1,000 species, a dedicated CNN is assigned to each cluster, enabling cluster-specific feature learning. During inference, each test image passes through the same ResNet50 extractor, is assigned to the nearest cluster by the trained k-means model and is classified by the corresponding CNN. A majority-voting mechanism consolidates predictions across multiple views of the same specimen to improve prediction stability. Figure 2 presents the complete end-to-end workflow, from raw image input to final species identification.

Feature Extraction using Resnet50

Figure 1 outlines the preprocessing and feature extraction workflow applied to the PlantCLEF 2015 repository. Before feature extraction, all images undergo three preprocessing steps:

(a) quality auditing to remove blurred, corrupted, or incomplete samples;

(b) pixel-intensity normalisation, scaling values to the [0, 1] range; and

(c) spatial resizing to 224 × 224 pixels to satisfy the ResNet50 input specification.

These operations ensure consistent image quality and reduce training instability caused by variations in image acquisition conditions.

Feature extraction is performed using ResNet50, a 50-layer deep residual network pre-trained on the

ImageNet dataset. ResNet50 was selected because residual learning architectures have demonstrated strong capability in capturing fine-grained visual characteristics while maintaining stable optimisation in deep networks. In plant species classification, such hierarchical representations are particularly important for identifying subtle botanical structures, including venation patterns, leaf margins, texture variations and surface morphology (17-19). The defining characteristic of ResNet50 is its residual connection mechanism, which allows gradients to propagate directly through earlier network layers during backpropagation. This design mitigates the vanishing gradient problem commonly encountered in very deep neural architectures and enables stable learning of hierarchical feature representations. Lower layers encode low-level visual information such as edges and colour gradients, whereas deeper layers capture increasingly complex botanical structures relevant to species discrimination.

For this study, the original classification head of ResNet50 is removed so that the network functions exclusively as a feature extractor within a transfer-learning framework. The Global Average Pooling (GAP) layer generates a 2,048-dimensional feature vector for each image. This compact feature representation preserves discriminative morphological characteristics necessary for distinguishing closely related plant species while reducing the original image data to a dimensionality suitable for clustering and downstream classification. Figure 1 presents the feature extraction and clustering workflow of the proposed framework.

Clustering for Feature Grouping

Following feature extraction, k-means clustering is applied to partition the feature space into structurally coherent groups based on Euclidean distance between feature vectors and cluster centroids. K-means was selected because of its computational efficiency, scalability and effectiveness in grouping high-dimensional feature representations according to visual similarity. In the context of plant species identification, this clustering stage reduces dataset complexity by organising morphologically similar samples into more homogeneous subsets prior to classification. This intermediate grouping strategy improves feature-space separability and reduces inter-class confusion, allowing each downstream CNN classifier to focus on a narrower distribution of plant species (20-22). Consequently, the cluster-specific CNNs can learn more specialised feature representations than a single global classifier trained across all species simultaneously.

The optimal number of clusters was determined using Silhouette Score analysis, which evaluates cluster quality by measuring the similarity of samples within the same cluster relative to neighbouring clusters. Silhouette values range from −1 to +1, where higher values indicate improved intra-cluster cohesion and inter-cluster separation. Cluster configurations ranging from k=5 to k=25 was systematically evaluated using Silhouette Score analysis. The evaluated configurations produced scores of 0.512 [k=5], 0.589 [k=10], 0.684 [k=15], 0.610 [k=20] and 0.545 [k=25]. Among the tested configurations, k=15 achieved the highest Silhouette Score, indicating the most effective balance between intra-cluster cohesion and inter-cluster separation within the PlantCLEF 2015 feature space. This configuration was therefore adopted for all subsequent experiments.

CNNs for Species Classification

The final stage employs convolutional neural networks (CNNs) to map extracted feature descriptors to plant species labels. CNNs learn hierarchical representations through successive convolutional operations (14), where early layers detect low-level visual features, intermediate layers identify structural patterns such as leaf contours and venation and deeper layers encode species-discriminative botanical traits, thereby reducing misclassification risk (19). As illustrated in Figure 2, each of the 15 clusters is assigned a dedicated CNN classifier, enabling each model to learn feature distributions specific to the species within its cluster (23-25).

The CNN architecture accepts a 224 × 224 RGB input and consists of three convolutional blocks. Preliminary experiments indicated that the three-layer configuration provided an effective balance between classification performance and computational complexity for the cluster-level classification task. The third convolutional layer improved hierarchical feature representation compared with shallower alternatives, whereas deeper configurations did not provide consistent performance gains within the cluster-constrained feature space.

Figure 3 presents the schematic architecture of the proposed cluster-specific CNN classifier, illustrating the sequential convolutional, pooling and fully connected stages used for plant species classification.

The first convolutional block applies 32 filters with ReLU activation and max-pooling, producing a 111 × 111 × 32 feature map that captures edges and surface textures (26). A second block with 64 filters reduces the spatial output to 54 × 54 × 64, encoding intermediate structural information such as leaf contours and vein networks. The third block employs 128 filters, compressing the representation to 26 × 26 × 128 while preserving discriminative botanical characteristics through progressive pooling operations (26).

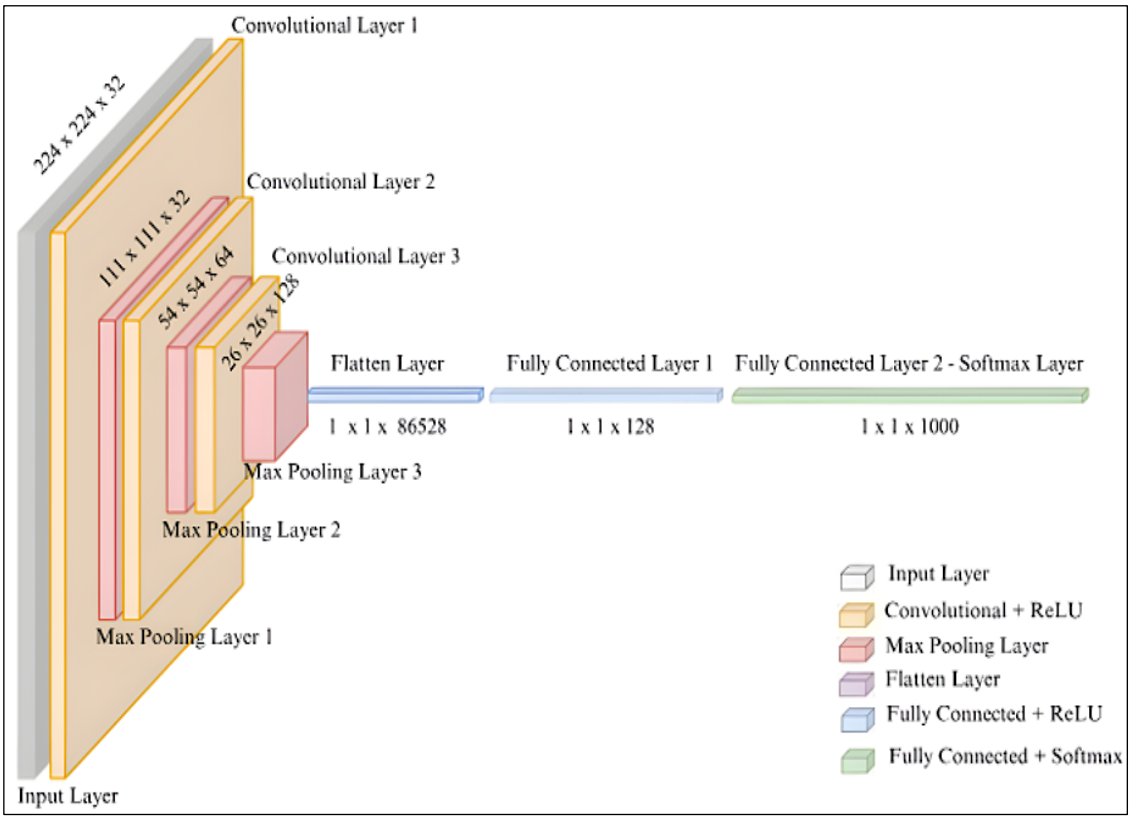

Figure 3: Schematic Representation of the Custom CNN for Plant Species Classification

Figure 4(A-L) illustrates the transition from low-level texture extraction in earlier layers to high-level discriminative feature encoding in deeper layers, thereby demonstrating the hierarchical abstract-tion behaviour learned by the proposed CNN architecture.

After the final pooling stage, the feature maps are flattened and passed to a fully connected layer containing 128 neurons, followed by a dropout layer with a rate of 0.5 for regularisation. A final output layer consisting of 1,000 neurons with Softmax activation converts network logits into probability scores for species classification.

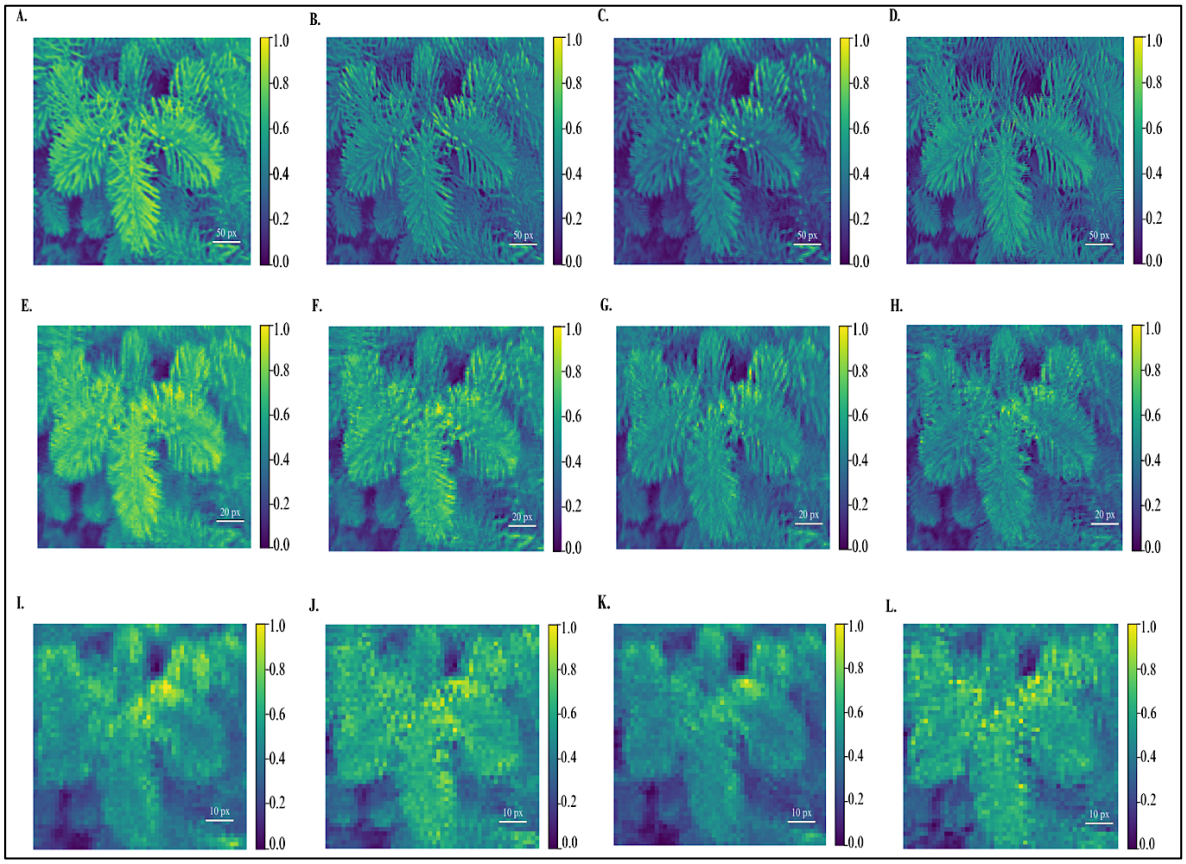

Figure 4: Visualizing the Internal Pattern Recognition and Feature Activations across Successive Convolutional Filtering Layers. (A-D) Convolutional Layer 1: 32 Filters, (A) Filter 16, Mean Act: 0.448, (B) Filter 15, Mean Act: 0.229, (C) Filter 11, Mean Act: 0.221, (D) Filter 8, Mean Act: 0.198; (E-H) Convolutional Layer 2: 64 Filters, (E) Filter 56, Mean Act: 0.322, (F) Filter 57, Mean Act: 0.197, (G) Filter 4, Mean Act: 0.174, (H) Filter 16, Mean Act: 0.148; (I-L) Convolutional Layer 3: 128 Filters, (I) Filter 9, Mean Act: 0.148, (J) Filter 72, Mean Act: 0.146, (K) Filter 120, Mean Act: 0.137, (L) Filter 5, Mean Act: 0.115

Table 1: Architectural Specifications of the Custom CNN Framework

Functional Block (Type)	Filters / Neurons	Filter	Stride	Output Shape	Parameters
Input	–	–	–	224 × 224 × 3	0
Conv2D_1 (ReLU)	32	[3, 3]	[1, 1]	222 × 222 × 32	896
MaxPooling2D_1	–	[2, 2]	[2, 2]	111 × 111 × 32	0
Conv2D_2 (ReLU)	64	[3, 3]	[1, 1]	109 × 109 × 64	18,496
MaxPooling2D_2	–	[2, 2]	[2, 2]	54 × 54 × 64	0
Conv2D_3 (ReLU)	128	[3, 3]	[1, 1]	52 × 52 × 128	73,856
MaxPooling2D_3	–	[2, 2]	[2, 2]	26 × 26 × 128	0
Flatten	–	–	–	86,528	0
Dense_1(ReLU)	128	–	–	128	11,075,712
Dropout [0.5]	–	–	–	128	0
Dense (softmax output layer)	1000	–	–	1000	129,000
Total Trainable Params:				11,297,960 (43.10 MB)	

Note: The proposed CNN architecture consists of three convolutional blocks followed by fully connected layers for high-dimensional feature learning and plant species classification. The framework contains a total of 11,297,960 trainable parameters occupying 43.10 MB memory space.

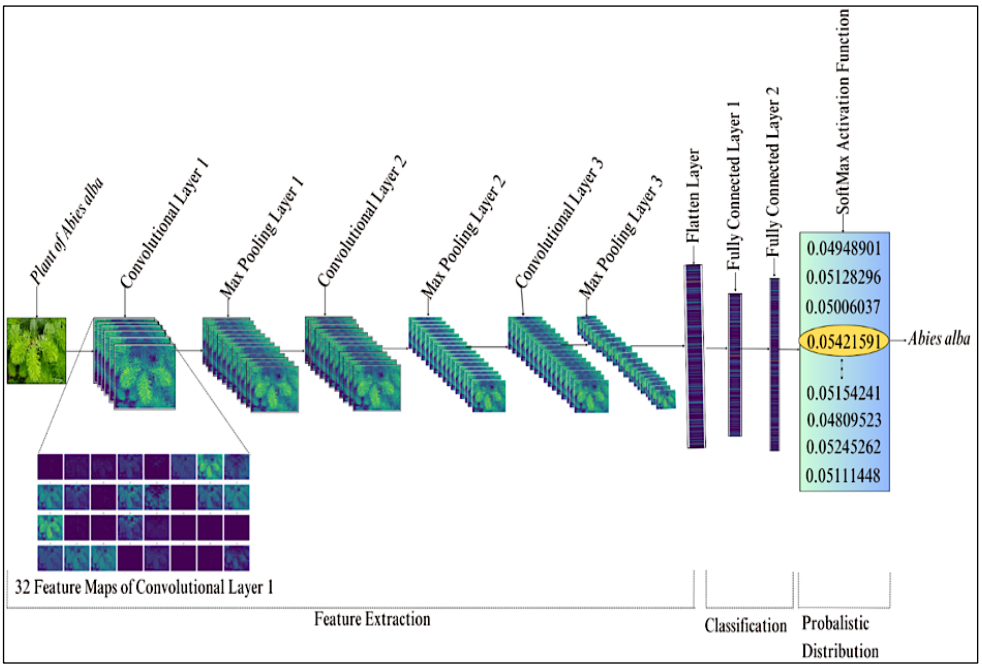

Figure 5: Illustration of Feature Map Transformation on Each Layer of the Proposed CNN as a Classifier

To mitigate overfitting in clusters with limited training samples, early stopping and within-cluster data augmentation were incorporated during training. Data augmentation included random horizontal flipping, rotation up to ±20° and brightness adjustment within a ±15% range. These regularisation strategies improved model generalisation and reduced the risk of cluster-specific overfitting. To improve prediction stability, a majority-voting mechanism is applied during inference. Each available view of a plant specimen is classified independently and the species receiving the highest number of predictions is assigned as the final label (27, 28). This strategy reduces the influence of image quality variation and partial occlusion of plant organs. Figure 5 illustrates the hierarchical transformation of visual features from raw input patterns to high-level species-discriminative representations across the proposed CNN architecture, demonstrating how successive convolutional stages progressively refine feature abstraction for fine-grained plant classification. The complete architectural specifications are summarised in Table 1.

Experimental Setup

The computational phase of this study was conducted on a high-tier infrastructure equipped with an NVIDIA T4 Graphics Processing Unit (GPU). This hardware features 2,560 CUDA cores and 320 Tensor Cores, supported by 16 GB of GDDR6 memory with a high-speed bandwidth of [320 GB/s]. Such a configuration ensures the necessary processing power for demanding deep learning operations. The host environment used an Intel Xeon processor and [51GB] of system memory, running on an Ubuntu-based platform optimized with CUDA 11.x libraries to enable accelerated GPU performance. For developing and training CNN models, industry-standard deep learning frameworks, including TensorFlow 2.x and PyTorch 1.x/2.x, were used. This hardware-software configuration ensured efficient execution of training, feature extraction and model inference, enabling timely and reliable experimental evaluation.

Dataset Used

The empirical foundation of this study is the PlantCLEF 2015 dataset, comprising 113,205 visual records spanning 1,000 plant species, including ferns, herbs and trees native to Western Europe. The dataset was assembled through a multi-year citizen science initiative led by Tela Botanica, with contributions from professional and amateur botanists (29). Images were captured under diverse climatic conditions, seasons and imaging environments, introducing substantial natural variability that makes the dataset a realistic benchmark for automated plant species identification. The repository includes multiple plant organs and viewpoints, including leaf scans, stems, reproductive structures (flowers and

fruits), branches and full-plant views (29, 30). Metadata such as geographic coordinates, acquisition timestamps, EXIF information and community-based image quality ratings from the Pictoflora platform are also provided.

Following preprocessing and quality assessment, 108,412 images were retained for model development and evaluation, while blurred, corrupted, or structurally incomplete samples were excluded. To address class imbalance, within-cluster data augmentation, including random horizontal flipping, rotation up to ±20° and brightness adjustment within a ±15% range, was applied to underrepresented species. In addition, class-weighted loss functions were incorporated during CNN training to assign higher penalty weights to minority classes, thereby improving balanced classification performance across both common and rare species categories.

System Evaluation Criteria

To rigorously quantify the classification efficacy, the following fundamental parameters are utilized:

(a) Correct Identifications (TP): The count of target varieties accurately categorized.

(b) Correct Rejections (TN): Instances where non-target species were successfully excluded.

(c) Incorrect Assignments (FP): Cases where non-target samples were erroneously designated as the target species.

(d) Missed Identifications (FN): Target specimens that the system failed to recognize correctly.

Based on these primary variables, the following performance indicators are calculated to provide a comprehensive analysis of the system's botanical identification success:

Overall Accuracy This represents the proportion of total correct predictions relative to the entire dataset. The formula expresses as Equation [1]:

$$\text{Accuracy} = (TP + TN) / TP + TN + FP + FN \quad [1]$$

a) Predictive Precision: This measure evaluates the dependability of the model's positive assignments by comparing successful hits to the sum of all positive claims (TP + FP).

b) Sensitivity (Recall): Recall measures the model's proficiency in detecting all relevant specimens, calculated as the proportion of original positive cases successfully identified (TP + FN).

c) F1-Measure: The F1-score acts as a balanced average of precision and sensitivity. It serves as a critical metric for judging the model's stability, particularly when balancing the risks of false triggers versus overlooked samples.

Results and Discussion

This chapter presents the analytical results and systematic benchmarking of the architectures using structured data and visual plots. The findings highlight the framework's effectiveness in handling botanical identification challenges. To ensure statistical reliability, a 5-fold cross-validation protocol was applied, dividing the PlantCLEF 2015 dataset into training [60%], validation [20%] and testing [20%] subsets. After optimization on the training set, models were evaluated on unseen test data to measure generalization performance. For rigorous comparison, established neural networks such as ResNet50 (12), InceptionV3 (31), Xception (32), EfficientNet (33) and MobileNet (34) were assessed alongside the proposed CNN under a unified computational environment to ensure fair and unbiased evaluation.

Performance of Models

For systematic assessment, five to six instances were selected from each category, yielding 5,549 evaluation samples organized by clustering four to six images per species from the test set [20%] of the PlantCLEF 2015 repository. By comparing ground-truth labels with predicted assignments, a confusion matrix was generated for each architecture.

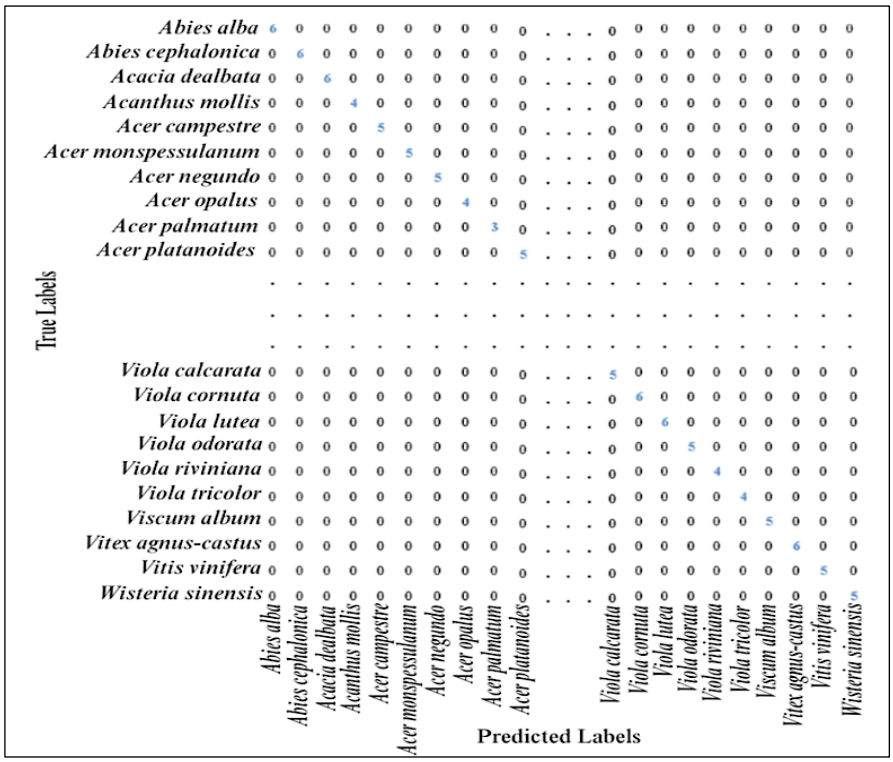

Figure 6: Confusion Matrix for Resnet50 as a Classifier

This matrix provides an in-depth analysis of classification performance, revealing both accuracy and error patterns. A sample confusion matrix for ResNet50 across 1,000 botanical species is shown in Figure 6. In this visualization, horizontal axes represent ground-truth labels, while vertical axes denote predicted labels. Each cell captures the count or percentage of instances assigned to a particular true-versus-predicted pair. Values along the main diagonal indicate correct classifications, while off-diagonal elements highlight misclassifications. Ideally, most data should concentrate on the main diagonal; non-zero off-diagonal entries reflect classification errors.

Quantitative data extracted from confusion matrices enable the calculation of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). These metrics are used to compute performance indicators, including precision, recall (sensitivity) and F1-score for all tested frameworks. Figure 7 shows the confusion matrix for the proposed CNN architecture across 1,000 species, effectively highlighting the classifier's strengths while identifying species pairs that contribute to classification errors. Table 2 further summarizes the performance of ResNet50 over the 1,000-class dataset.

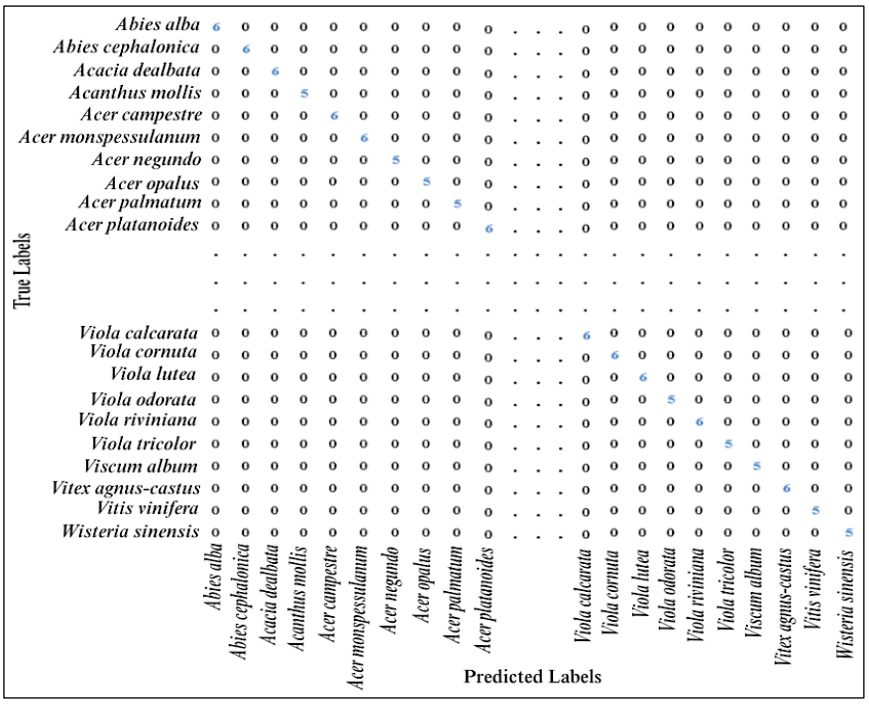

Figure 7: Confusion Matrix for the Proposed CNN as a Classifier

Table 2: Classification Report for ResNet50 as a Classifier

Class	TP	FP	TN	FN	Accuracy	Precision	Recall	F1 Score
Abies alba	6	0	5543	0	1.00	1.00	1.00	1.00
Abies cephalonica	6	2	5541	0	0.99	0.75	1.00	0.85
Acacia dealbata	6	0	5543	0	1.00	1.00	1.00	1.00
Acanthus mollis	4	0	5543	2	0.99	1.00	0.66	0.80
Acer campestre	5	1	5542	1	0.99	0.83	0.83	0.83
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
Viola tricolor	4	1	5543	1	0.99	0.80	0.80	0.80
Viscum album	5	0	5544	0	1.00	1.00	1.00	1.00
Vitex agnus-castus	6	0	5543	0	1.00	1.00	1.00	1.00
Vitis vinifera	5	1	5543	0	0.99	0.83	1.00	0.90
Wisteria sinensis	5	1	5543	0	0.99	0.83	1.00	0.90

Note: TP, FP, TN and FN denote True Positive, False Positive, True Negative and False Negative, respectively. Performance metrics are reported for each plant species class. Only representative class-wise results are presented due to space limitations.

Table 3: Classification Report for the Proposed CNN as a Classifier

Class	TP	FP	TN	FN	Accuracy	Precision	Recall	F1 Score
Abies alba	6	0	5543	0	1.00	1.00	1.00	1.00
Abies cephalonica	6	0	5543	0	1.00	1.00	1.00	1.00
Acacia dealbata	6	0	5543	0	1.00	1.00	1.00	1.00
Acanthus mollis	5	0	5543	1	0.99	1.00	0.83	0.90
Acer campestre	6	1	5543	0	1.00	1.00	1.00	1.00
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
Viola tricolor	5	0	5544	0	1.00	1.00	1.00	1.00
Viscum album	5	0	5544	0	1.00	1.00	1.00	1.00
Vitex agnus-castus	6	0	5543	0	1.00	1.00	1.00	1.00
Vitis vinifera	5	0	5544	0	1.00	1.00	1.00	1.00
Wisteria sinensis	5	0	5544	0	1.00	1.00	1.00	1.00

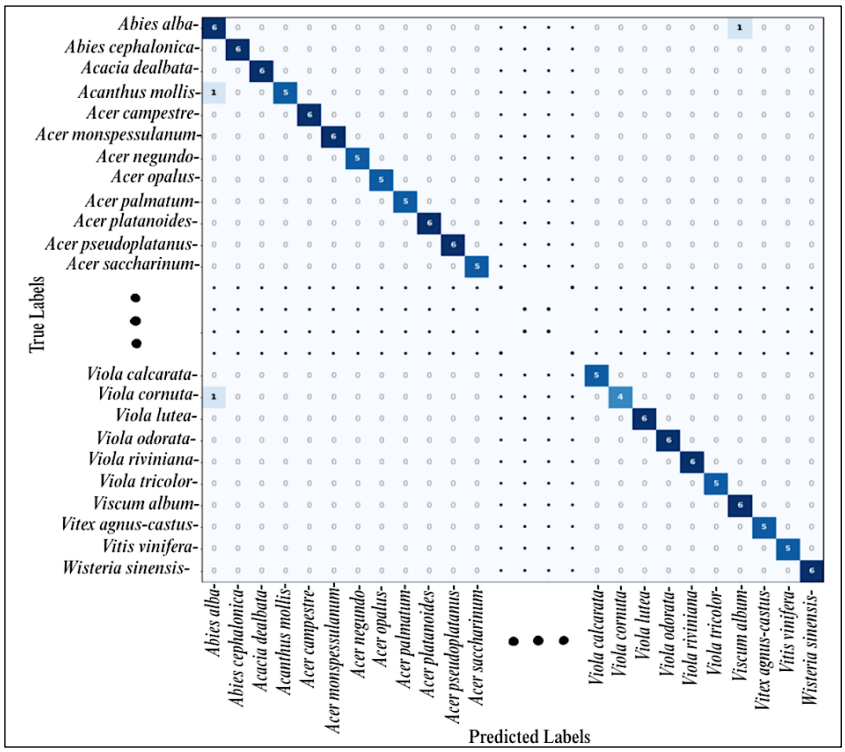

Figure 8: Heat Map Representation of the Proposed CNN as a Classifier

Most predictions fall along the diagonal, indicating correct classifications. These results are largely attributed to the improved discriminative capability achieved through cluster-specific CNN selection from among the 15 classifiers determined via k-means feature grouping, consistent with findings from hierarchical plant classification systems that demonstrate performance gains when models are specialized to specific morphological subgroups (29, 30). Figure 8 illustrates the classification outcome chart for the proposed CNN, color-coded by gradient intensity. This visualization makes it easy to identify where the system performed optimally and where minor classification overlaps occurred.

The behavior of validation curves reflects the model's generalization capability and training stability. Over 150 epochs (illustrated in Figures 9A and 9B, both training and validation loss gradually decreased while accuracy showed a consistent upward trend. However, slight oscillations in validation metrics indicate potential sensitivity to training dynamics, potentially suggesting mild overfitting. Such variations can arise from learning rate configuration, data distribution imbalance, or regularization mechanisms (26). To ensure comprehensive performance evaluation, additional indicators-precision, recall (sensitivity) and F1-score-were incorporated into the analysis. These complementary metrics provide deeper insight into classification reliability beyond overall accuracy. The experimental results demonstrate strong predictive performance across the majority of plant species. Although limited false-positive and false-negative predictions were observed in specific classes, overall classification effectiveness remains robust.

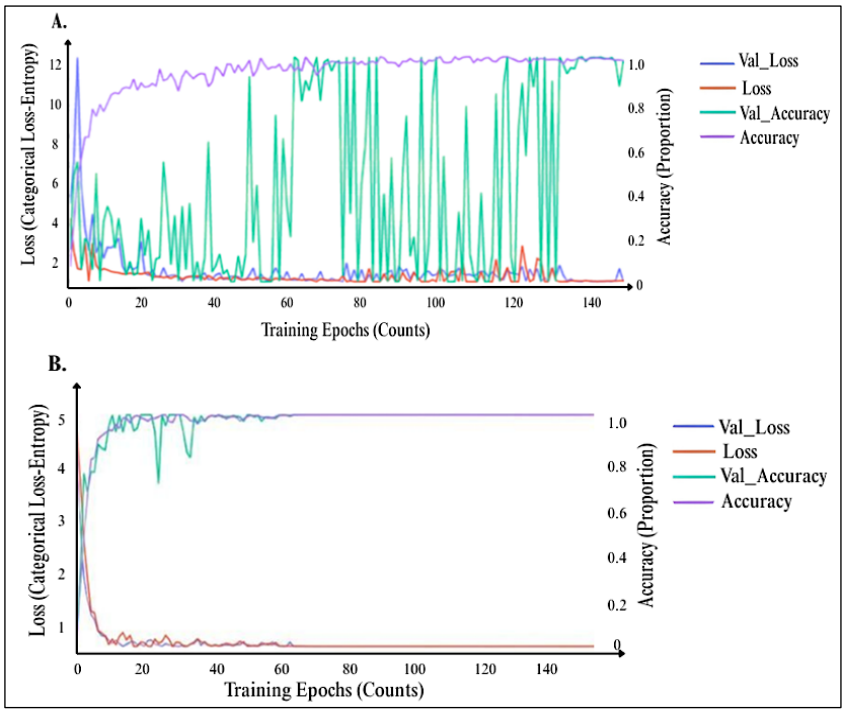

Figure 9: (A) Training Results for The Resnet50 Classifier Showing Loss and Accuracy, (B) Training Progress of the Proposed CNN Framework Showing Error and Success Rates

Table 3 presents detailed classification results for all 1,000 species. Most classes achieve very high precision, recall and F1-scores, demonstrating strong discriminative ability. A few species, such as *Anemone coronaria*, record lower F1-scores [e.g., 0.73], likely due to visual similarity with related species. This pattern aligns with observations in recent fine-grained classification studies, where morphologically similar species consistently present greater identification challenges (16, 29). Overall, the model performs reliably across the dataset. Performance metrics based on TP, TN, FP and FN are summarized in Table 4. The proposed model achieves the highest accuracy [0.9800], outperforming all baseline networks. Among the compared models, Xception [0.8839] slightly surpasses InceptionV3 [0.8824] and EfficientNet [0.8820], while ResNet50 [0.8649] and MobileNet [0.8773] show lower accuracy (12, 31-34). The proposed framework also reports the lowest loss [0.0615], indicating better generalization than ResNet50 [0.1351] and EfficientNet [0.1180].

As shown in Figure 10, the proposed system consistently achieves higher precision, recall and F1-score, ensuring balanced and reliable classification. In terms of efficiency, the model requires 61.5 seconds per epoch for training. Although MobileNet is faster [50.5s], it achieves lower precision. EfficientNet [241.2s] and InceptionV3 [208.5s] require significantly more training time (31, 33, 34), as illustrated in Figure 10(E). Additionally, the model achieves an average inference time of 18 ms per image on standard hardware (NVIDIA Tesla T4 GPU), making it suitable for real-time field applications in biodiversity monitoring and precision agriculture. According to Figure 10(F), the proposed architecture maintains a moderate parameter count compared to larger models such as InceptionV3 and EfficientNet. This balanced design enables deployment in resource-limited environments without sacrificing accuracy.

Visual Representation of Metrics

Figure 10 presents a detailed graphical analysis of primary evaluation criteria, including classification accuracy, precision and recall. Furthermore, it provides a comparative assessment of F1-score, average training duration per epoch and total model parameters. The results reveal several notable trends: (a) The proposed model achieved substantial accuracy gains compared with conventional baseline models, (b) Execution time-accuracy trade-offs are observed in lightweight architectures such as MobileNet, where reduced computation comes at the cost of lower performance, (c) The proposed model demonstrates balanced efficiency, maintaining

high classification accuracy while reducing computational overhead.

Error Analysis

The results in Table 4 and Figure 10 demonstrate that the proposed architecture outperforms conventional models, including ResNet-50, Inception-V3, Xception, EfficientNet and MobileNet. Although minor misclassifications occurred among visually similar species, performance can be further improved by expanding the training data, adopting advanced augmentation strategies, or incorporating attention mechanisms (6, 11). Similar trends have been observed in recent plant classification studies, where deep architectures consistently achieve higher accuracy on fine-grained botanical datasets compared to traditional machine learning approaches (9-11). As illustrated in Figure 10(A-D), the proposed model achieves the highest accuracy, precision, F1-score and recall, confirming both reliability and balanced predictive capability. Figure 10E shows competitive training time per epoch, indicating reasonable computational efficiency. In contrast, Figure 10F demonstrates that strong performance is maintained despite a relatively compact parameter size compared to heavier architectures such as EfficientNet.

The model's effectiveness stems from combining high accuracy with reduced architectural complexity. Using ResNet-50 as the feature extraction backbone, hierarchical representations are effectively captured, while residual connections mitigate vanishing gradients and preserve low-level visual information. This design enables robust identification of subtle botanical traits essential for distinguishing visually similar plant species, consistent with findings that reported that residual architectures excel at capturing fine-grained morphological variations in plant leaves (12, 17, 18).

Table 4: Comparative Performance of Traditional Classifiers and Proposed CNN Classifier

Classifier	Avg. Accuracy	Avg. Loss	Avg. Time (s)
ResNet50	0.8649	0.1351	196.5
InceptionV3	0.8824	0.1176	208.5
Xception	0.8839	0.1161	167.3
EfficientNet	0.8820	0.1180	241.2
MobileNet	0.8773	0.1227	50.5
Proposed CNN	**0.9800**	**0.0615**	**61.5**

Note: Higher accuracy and lower loss indicate better classification performance, while lower execution time indicates better efficiency. Bold values represent the best results.

To enhance the model's discriminative capabilities, k-means clustering organizes the extracted features into coherent groups corresponding to visual similarity patterns. This unsupervised partitioning of the high-dimensional feature space creates distinct, compact clusters that minimize intra-class variance and maximize inter-class separation, an approach supported by recent work in hierarchical plant classification systems (24, 25).

The final classification stage operates on this refined feature set, reducing error rates through hierarchical processing. This multi-stage approach offers flexibility to fine-tune the CNN for specific environmental or categorical conditions. The architecture excels at hierarchical learning, transitioning from elementary edges to sophisticated semantic structures-particularly critical in botany, where minor variations in shape and texture determine species identity. By incorporating ReLU activation and max pooling, the model enhances robustness while maintaining computational efficiency. The incremental increase in convolutional filters allows the network to capture both fine details and broad, abstract patterns. This scalable design ensures the model remains effective for plant identification while being adaptable to broader image classification challenges in real-world, resource-constrained environments.

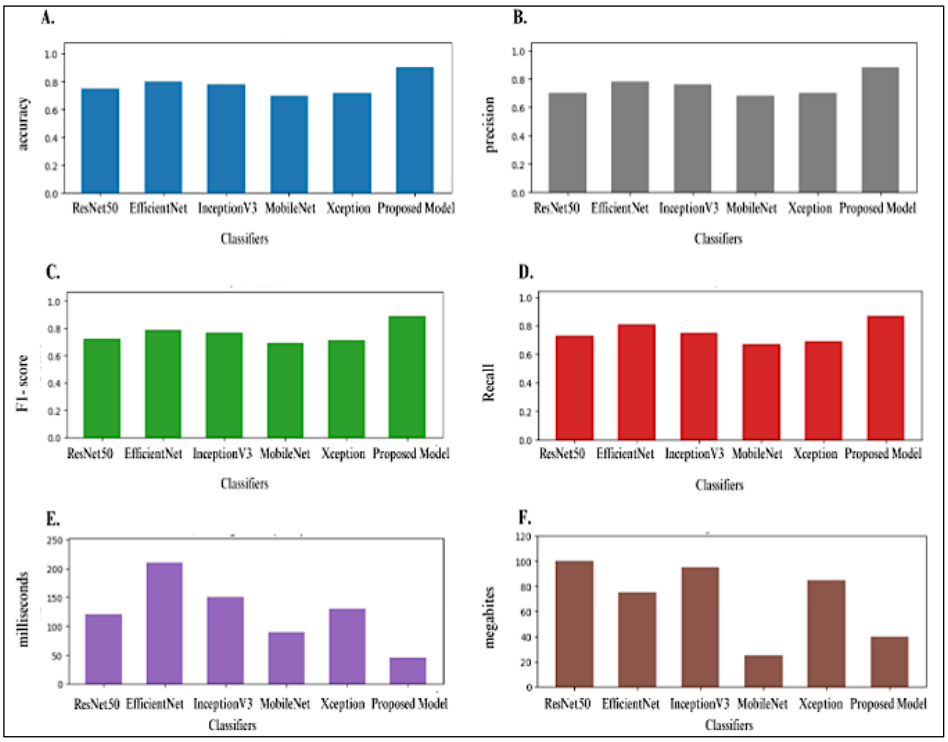

Figure 10: Comparative Performance of Deep Learning Classifiers. (A) Accuracy, (B) Precision, (C) F1-Score, (D) Recall, (E) Average Time Per Epoch, (F) Average Parameter Size

Conclusion

This study proposed a hierarchical plant identification framework integrating ResNet50 feature extraction, k-means clustering and 15 cluster-specific CNN classifiers for fine-grained plant species recognition. Evaluated on the PlantCLEF 2015 dataset containing 1,000 species and 113,205 images, the framework achieved [0.9800] classification accuracy with [0.0615] loss under five-fold cross-validation. The proposed approach outperformed all evaluated baselines and delivered a [0.1151] accuracy improvement over standalone ResNet50, demonstrating the effectiveness of cluster-guided classification for reducing inter-species similarity challenges. Furthermore, an inference latency of [18ms] and an average epoch time of [61.5s] confirm the practical feasibility of the framework for real-world deployment.

Limitations

This study remains limited by single-dataset evaluation, increased architectural complexity due to multiple independent CNN classifiers and the absence of formal statistical significance analysis.

Future Research Directions

Future work will focus on cross-dataset generalization, lightweight attention-enhanced architectures for edge and mobile environments and broader applications in crop disease identification and biodiversity monitoring.

Abbreviations

CNN: Convolutional Neural Network, DL: Deep Learning, EXIF: Exchangeable Image File Format, FC: Fully Connected (Layer), GAP: Global Average Pooling, ILSVRC: ImageNet Large Scale Visual Recognition Challenge, ReLU: Rectified Linear Unit, RGB - Red, Green, Blue (Color Model), ResNet: Residual Network, TL: Transfer Learning.

Acknowledgment

The authors gratefully acknowledge the support and facilities provided by Annamalai University for the successful completion of this research work.

Author Contributions

All authors contributed equally to the conceptualization, methodology, experimentation and preparation of the manuscript.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data Availability

The data used in this study are available from the corresponding author upon reasonable request.

Declaration of Artificial Intelligence (AI) Assistance Process

The authors used generative AI tools only for language improvement and rephrasing purposes. The scientific content, methodology, results and conclusions are entirely the original work of the authors. The authors take full responsibility for the content's originality, interpretation and accuracy.

Ethics Approval

Not applicable.

Funding

This research received no external funding from any public, commercial, or non-profit funding agencies.

References

1. Carranza-Rojas J, Goeau H, Bonnet P, *et al*. Going deeper in the automated identification of herbarium specimens. BMC Evol Biol. 2017;17(1):181. doi:10.1186/s12862-017-1014-z
2. Alsakar YM, Sakr NA, Elmogy M. An enhanced classification system of various rice plant diseases based on a multi-level handcrafted feature extraction technique. Sci Rep. 2024;14(1):30601. doi: 10.1038/s41598-024-81143-1
3. Zaman W, Ayaz A, Park S. Integrating morphological and molecular data in plant taxonomy. Pak J Bot. 2025;57(4):1453-1466. doi: 10.30848/PJB2025-4(27)
4. Sujatha R, Krishnan S, Chatterjee JM, *et al*. Advancing plant leaf disease detection by integrating machine learning and deep learning. Sci Rep. 2025;15(1):11552. doi: 10.1038/s41598-024-72197-2
5. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations (ICLR); 2015 May; San Diego, CA. arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556
6. Dubey RK, Choubey DK. Adaptive feature selection with a deep learning MBi-LSTM model-based paddy plant leaf disease classification. Multimed Tools Appl. 2024;83(9):1. doi: 10.1007/s11042-023-16475-7
7. Shi Y, Wang T, Skidmore AK, *et al*. Important LiDAR metrics for discriminating forest tree species in Central Europe. ISPRS J Photogramm Remote Sens. 2018; 137:163-174. doi: 10.1016/j.isprsjprs.2018.02.002-29.
8. Tan JW, Chang SW, Abdul-Kareem S, *et al*. Deep learning for plant species classification using leaf vein morphometrics. IEEE/ACM Trans Comput Biol Bioinform. 2020;17(1):82-90. doi: 10.1109/TCBB.2018.2848653
9. Meselhy E, Gamal F. A comparative study of state-of-the-art algorithms for plant recognition and classification on a large dataset. Int J Comput Inf. 2023;10(3):164-174. doi: 10.21608/ijci.2023.236584.1144
10. Balafas V, Karantoumanis E, Louta M, *et al*. Machine learning and deep learning for plant disease classification and detection. IEEE Access. 2023; 11:114352-77. doi: 10.1109/ACCESS.2023.3324722
11. Islam MT, Rahman W, Hossain MS, *et al.* Medicinal plant classification using particle swarm optimized cascaded network. IEEE Access. 2024; 12:42465-78. doi: 10.1109/ACCESS.2024.3378262
12. He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR);2016;770-8. doi: 10.1109/CVPR.2016.90
13. Varadharajan PK, Rajakumar R. Performance analysis of a convolutional neural network in image classification. Int J Comput Appl. 2023;184(48):14-8. doi: 10.5120/ijca2023922597
14. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44. doi: 10.1038/nature14539
15. Saleem G, Akhtar M, Ahmed N, *et al*. Automated analysis of visual leaf shape features for plant classification. Comput Electron Agric. 2019; 157: 270-80. doi: 10.1016/j.compag.2018.12.038
16. Šulc M, Matas J. Fine-grained recognition of plants from images. Plant Methods. 2017;13(1):1-14. doi: 10.1186/s13007-017-0265-4
17. Palkar JD, Deshpande AS. Performance evaluation of CNN, ResNet50 and hybrid architectures for automated grape leaf disease detection. Appl Fruit Sci. 2026;68(2):98. doi: 10.1007/s10341-026-01813-4
18. Deshpande R, Patidar H. Detection of leaf disease in tomato plants using a lightweight parallel deep convolutional neural network. Arch Phytopathology Plant Prot. 2023;56(9):707-20. doi: 10.1080/03235408.2023.2216359
19. Grinblat GL, Uzal LC, Larese MG, *et al*. Deep learning for plant identification using vein morphological patterns. Comput Electron Agric. 2016; 127:418-24. doi:10.1016/j.compag.2016.07.003
20. Sinaga KP, Yang MS. Unsupervised K-means clustering algorithm. IEEE Access. 2020; 8:80716-80727. doi: 10.1109/ACCESS.2020.2988796
21. Wang W, He Y, Ma L, *et al*. Latent feature group learning for high-dimensional data clustering. Information. 2019;10(6):208. doi: 10.3390/info10060208
22. Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans Knowl Data Eng. 2005;17(4):491-502. doi: 10.1109/TKDE.2005.66
23. Zou X, Zhou L, Li K, *et al*. Multi-task cascade deep convolutional neural networks for large-scale commodity recognition. Neural Comput Appl. 2020;32(10):5633-5647. doi: 10.1007/s00521-019-04311-9
24. Lin Q, Zhao J, Du B, *et al*. MEDNet: Multiexpert detection network with unsupervised clustering of training samples. IEEE Trans Geosci Remote Sens. 2021; 60:1-14.

doi: 10.1109/TGRS.2021.3093556
25. Cai Z, Vasconcelos N. Cascade R-CNN: High quality object detection and instance segmentation. IEEE Trans Pattern Anal Mach Intell. 2021;43(5):1483-98. doi: 10.1109/TPAMI.2019.2956516
26. Zafar A, Aamir M, Mohd Nawi N, *et al.* A comparison of pooling methods for convolutional neural networks. Appl Sci. 2022;12(17):8643. doi: 10.3390/app12178643
27. van der Velde M, Goëau H, Bonnet P, *et al*. Pl@ntNet Crops: Merging citizen science observations and structured survey data to improve crop recognition. Environ Res Lett.2023;18(2):024014. https://doi.org/10.1088/1748-9326/acadf3?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle
28. Faska Z, Khrissi L, Mountasser I, *et al.* An efficient weighted majority voting ensemble machine learning classifier framework for image segmentation. Eng Technol Appl Sci Res. 2026;16(3):35220-37. https://doi.org/10.48084/etasr
29. Goëau H, Bonnet P, Joly A, *et al.* LifeCLEF Plant Identification Task 2015. In: Working Notes of CLEF 2015. CEUR Workshop Proceedings. 1-13. https://www.researchgate.net/publication/280953357_LifeCLEF_Plant_Identification_Task_2015
30. Keivani M, Mazloum J, Sedaghatfar E, *et al*. Automated analysis of leaf shape, texture and color features for plant classification. Trait Signal. 2020;37(1):17-28. doi: 10.18280/ts.370103
31. Szegedy C, Vanhoucke V, Ioffe S, *et al*. Rethinking the inception architecture for computer vision. IEEE. 2016;2818-26. doi: 10.1109/CVPR.2016.308
32. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE CVPR; 2017;1251-8. doi: 10.1109/CVPR.2017.195
33. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th ICML; 2019 Jun; Long Beach, CA. PMLR 97:6105-6114. doi: 10.48550/arXiv.1905.11946
34. Howard AG, Zhu M, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv. 2017. doi: 10.48550/arXiv.1704.04861

How to Cite: Karnan A, Ragupathy R. An Innovative Method for Plant Species Classification Utilizing Convolutional Neural Networks. Int Res J Multidiscip Scope. 2026;7(3):185-200.
DOI: 10.47857/irjms.2026.v07i03.09761