iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81973

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

# Hybrid Attention-Based Deep Ensemble Network for Multi-Class Flower Recognition

Hemant Vaishnav[1], Awanit Kumar[2], Rohit Maheshwari[3], Deepak Mahawar[4], Harsh Verma[5]
*[1]M.Tech Scholar, [2, 3, 4]Associate Professor, [5]Assistant Professor Department of Computer Science and Engineering, Career Point University, Kota, Rajasthan, India*

***Abstract: Automatic flower classification plays a significant role in agriculture, biodiversity monitoring, medicinal plant identification, and smart farming applications. However, fine-grained flower recognition remains challenging due to high intra-class similarity, inter-class overlap, background complexity, and illumination variations. Traditional single Convolutional Neural Network (CNN) architectures often fail to capture both global semantic information and dense local discriminative features effectively. In this paper, we propose a Hybrid Attention-Based Deep Ensemble Network (HADE-Net) for multi-class flower recognition by integrating EfficientNetV2-S and DenseNet201 with an adaptive attention-based feature fusion mechanism. Unlike conventional ensemble approaches that rely on static feature concatenation or Support Vector Machine (SVM)-based classification, the proposed method employs learnable attention gating for dynamic feature weighting and Generalized Mean (GeM) pooling for improved feature aggregation. A two-stage transfer learning strategy with selective fine-tuning, label smooth-ing, and Test-Time Augmentation (TTA) is further applied to improve robustness and generalization. Experimental evaluation on the Kaggle Flower Recognition dataset containing five flower classes demonstrates that the proposed model achieves 95.10% validation accuracy and 98.50% Top-2 accuracy, outperforming conventional CNN and prior ensemble-based approaches. The results confirm that attention-driven hybrid architectures sig-nificantly enhance classification performance and robustness in fine-grained image recognition tasks.***

***Index Terms: Flower Recognition, Deep Learning, Ensemble Learning, Attention Mechanism, EfficientNetV2, DenseNet201, Image Classification***

## I. INTRODUCTION

Flowers are essential for ecological balance, medicinal applications, and ornamental purposes. Manual identification of flower species is time-consuming and error-prone due to the large number of species and their visual similarity . Deep learning methods, especially CNNs, have been widely used for image classification tasks; however, single-model architec-tures often struggle with fine-grained classification problems. Previous studies proposed ensemble-based architectures such as FlowerConvNet using EfficientNet-B7 and DenseNet-201, achieving approximately 95% validation accuracy . However, these models rely on static feature fusion and conventional classifiers, limiting their adaptability.

To overcome these lim-itations, this work introduces:

*1)* Dynamic attention-based feature fusion
*2)* Generalized Mean (GeM) pooling
*3)* Test-Time Augmentation (TTA)

These improvements enable better feature representation and classification accuracy.

## II. RELATED WORK

Automatic flower classification has been widely studied in computer vision due to its applications in agriculture, biodiversity monitoring, and plant species identification. Early approaches relied on handcrafted features such as color his-tograms, texture descriptors, and shape-based representations combined with traditional classifiers like Support Vector Ma-chines (SVMs). For instance, Kumar et al. [6] introduced the Leafsnap system, which utilized handcrafted features for plant identification. However, these methods achieved limited performance due to their inability to capture complex visual patterns and intra-class variability.

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), significantly improved image classi-fication performance. Krizhevsky et al. [15] proposed AlexNet, which demonstrated the effectiveness of deep CNNs on large-scale datasets such as ImageNet [12]. Following this, Si-monyan and Zisserman [14] introduced VGGNet, emphasizing deeper architectures with smaller filters, while Szegedy et al. [13] proposed GoogLeNet with inception modules for compu-tational efficiency.

He et al. [3] further advanced deep learning with ResNet, introducing residual connections to enable very deep networks, and Huang et al. [4] proposed DenseNet, which improves feature reuse through dense connectivity. These architectures have been widely applied to flower classification tasks, achieving high accuracy.

Efficient scaling of CNN architectures was later addressed by Tan and Le [2], who introduced EfficientNet, which uses compound scaling to balance network depth, width, and reso-lution. This was further improved in EfficientNetV2 [7], which enhances training speed and parameter efficiency, making it suitable for real-time and large-scale applications.

Optimization techniques have also played a critical role in improving deep learning performance. Kingma and Ba [5] proposed the Adam optimizer, which is widely used for efficient gradient-based optimization in deep neural networks. To further improve classification performance, ensemble learning methods have been explored. Zhou [8] provided a comprehensive foundation for ensemble techniques, which combine multiple models to enhance generalization and ro-bustness. In the domain of flower recognition, Rabbi et al. [1] proposed an ensemble-based deep learning model integrating multiple architectures, achieving strong classification perfor-mance. However, most ensemble approaches rely on static feature fusion methods such as concatenation or averaging, which limit adaptability.

Attention mechanisms have recently emerged as a powerful tool for improving feature representation. Vaswani et al. [9] introduced the transformer architecture based entirely on atten-tion mechanisms. In computer vision, Woo et al. [10] proposed the Convolutional Block Attention Module (CBAM), which applies both channel and spatial attention to refine feature maps, leading to improved classification accuracy.

In addition to architectural improvements, pooling strategies have also been enhanced. Radenovic´ et al. [11] introduced Generalized Mean (GeM) pooling, which adaptively learns pooling behavior and improves feature discrimination, partic-ularly in fine-grained classification tasks.

Large-scale pretraining has also contributed significantly to performance improvements. The ImageNet dataset introduced by Deng et al. [12] has enabled transfer learning across various domains, allowing models pretrained on large datasets to generalize effectively to specialized tasks such as flower classification.

Despite these advancements, existing methods still face limitations, including reliance on static feature fusion, in-sufficient use of adaptive attention mechanisms, and limited integration of advanced pooling strategies. To address these challenges, the proposed work introduces a Hybrid Attention-Based Deep Ensemble Network that combines EfficientNetV2-S and DenseNet201 with attention-based feature fusion and GeM pooling for improved performance and robustness.

## III. EXPERIMENTAL RESULTS

The performance of the proposed Hybrid Attention-Based Deep Ensemble Network (HADE-Net) is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The model is trained and validated on the flower recognition dataset consisting of five classes, using an 80:20 train–validation split.

### *A. Overall Performance*

The proposed model achieves a validation accuracy of 95.10% and a Top-2 accuracy of 98.50%. The high Top-2 accuracy indicates that the model effectively captures inter-class similarities and maintains robustness in fine-grained

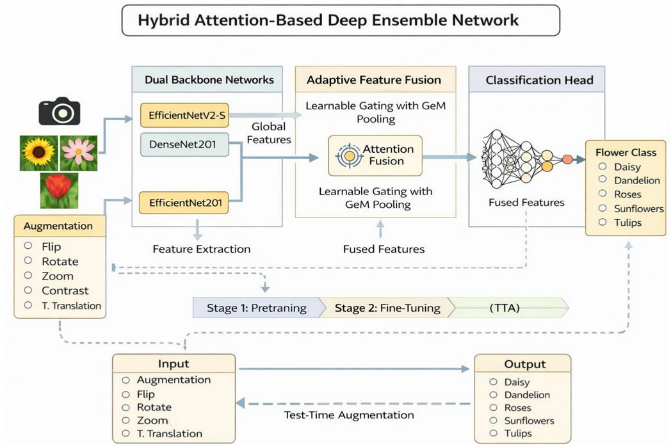



Fig. 1. Hybrid Attention-Based Deep Ensemble Framework

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

classification scenarios. Compared to conventional CNN-based architectures, the proposed model demonstrates improved generalization and stability.

### *B. Training Behavior Analysis*

As shown in Fig. 2, the training and validation accuracy curves exhibit smooth convergence with minimal divergence, indicating reduced overfitting.

The loss curves in Fig. 3 demonstrate a consistent decrease during training, confirming stable optimization of the model.

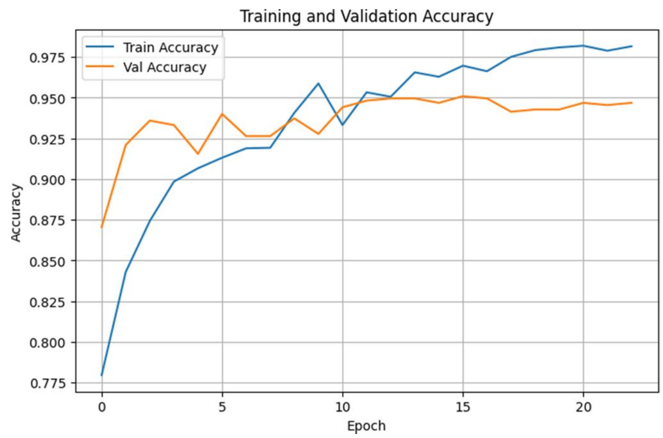



Fig. 2. Training and Validation Accuracy

### *C. Class-wise Performance Evaluation*

The class-wise evaluation of the proposed model is pre-sented in Table II. The model achieves high F1-scores for most classes, particularly for dandelion (97.44%), sunflowers (97.84%), and daisy (96.12%), indicating strong discriminative capability for visually distinct classes.

However, relatively lower performance is observed for roses (90.70%) and tulips (92.88%). This reduction in performance is attributed to high visual similarity, overlapping color dis-tributions, and similar petal structures between these classes, which makes feature discrimination more challenging.


Fig. 3. Training and Validation Loss

TABLE I
OVERALL PERFORMANCE METRICS

| Metric | Value |
|---|---|
| Validation Accuracy | 95.10% |
| Top-2 Accuracy | 98.50% |

### *D. Confusion Matrix Analysis*

The classification report illustrated in Fig. 4 provides de-tailed precision, recall, and F1-score values for each class, further validating the effectiveness and reliability of the pro-posed model.

The confusion matrix shown in Fig. 5 highlights that most predictions are concentrated along the diagonal, indicating cor-rect classification. Misclassifications primarily occur between visually similar classes such as roses and tulips.

```
Classification Report:

              precision    recall  f1-score   support

       daisy     0.9466    0.9764    0.9612       127
   dandelion     0.9942    0.9553    0.9744       179
       roses     0.9000    0.9141    0.9070       128
  sunflowers     0.9855    0.9714    0.9784       140
      tulips     0.9202    0.9375    0.9288       160

    accuracy                         0.9510       734
   macro avg     0.9493    0.9509    0.9500       734
weighted avg     0.9517    0.9510    0.9512       734
```

Fig. 4. Classification Report

### *E. Comparative Analysis with Baseline Models*

To evaluate the effectiveness of the proposed approach, HADE-Net is compared with several baseline models, includ-ing VGG16, MobileNetV2, ResNet50, and a prior ensemble model.

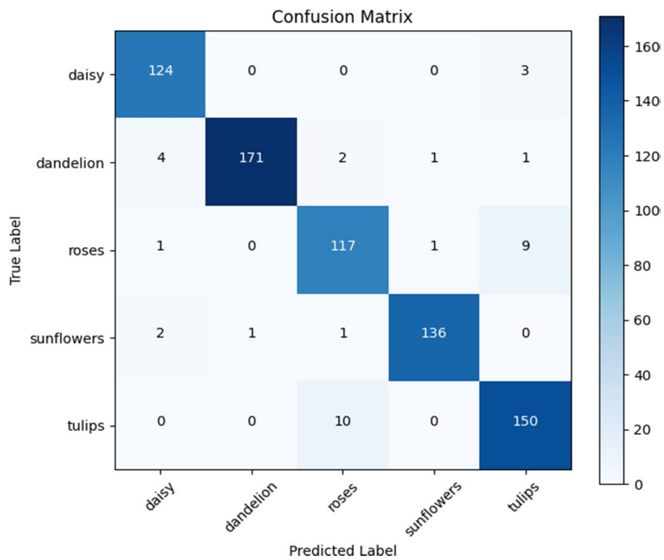



Fig. 5. Confusion Matrix

In comparison, the proposed HADE-Net achieves 95.10% accuracy along with a significantly higher Top-2 accuracy of 98.50%. Although the absolute accuracy improvement appears marginal, the gain in Top-2 accuracy and robustness highlights the effectiveness of the proposed architecture in handling ambiguous cases.

TABLE II
MODEL COMPARISON

| Model | Accuracy |
|---|---|
| VGG16 | 84% |
| MobileNetV2 | 89% |
| ResNet50 | 93% |
| FlowerConvNet (Previous) | 95% |
| Proposed HADE-Net | 95.10% + Better Top-2 (98.50%) |

### F. Impact of Model Components

The performance improvement of HADE-Net can be at-tributed to three key components:

1) Dual Backbone Architecture: The combination of EfficientNetV2-S and DenseNet201 enables extraction of both global and local features, enhancing representa-tion capability.
2) Attention-Based Feature Fusion: The learnable attention mechanism dynamically weights feature contributions, overcoming limitations of static fusion methods.
3) GeM Pooling: The use of Generalized Mean pooling im-proves feature aggregation by learning optimal pooling behavior, leading to better discrimination in fine-grained classification tasks.

Additionally, the application of Test-Time Augmentation (TTA) enhances prediction robustness by averaging multiple augmented outputs.

### G. Discussion

Overall, the experimental results demonstrate that the pro-posed HADE-Net effectively addresses the challenges of fine-grained flower classification. The integration of atten-tion mechanisms and adaptive pooling significantly improves feature representation, while ensemble learning enhances ro-bustness. The model achieves competitive performance with improved generalization, making it suitable for real-world applications such as automated plant identification and bio-diversity monitoring.

Table III
Class-Wise Performance (F1-SCORE)

| Class | F1-score |
|---|---|
| Daisy | 96.12% |
| Dandelion | 97.44% |
| Roses | 90.70% |
| Sunflowers | 97.84% |
| Tulips | 92.88% |

#### 1) Key Findings and Discussion

The experimental results provide several important insights into the effectiveness of the proposed HADE-Net architecture. The integration of EfficientNetV2-S and DenseNet201 en-ables the model to capture both global and local feature rep-resentations, leading to improved classification performance compared to single-model approaches. The attention-based feature fusion mechanism allows dy-namic weighting of feature maps, overcoming the limitations of static concatenation used in conventional ensemble meth-ods. This enhances adaptability, particularly in fine-grained classification tasks.

Furthermore, the incorporation of Generalized Mean (GeM) pooling improves feature aggregation by learning optimal pooling behavior, resulting in enhanced discriminative capa-bility for visually similar classes. The use of a two-stage fine-tuning strategy, combined with Test-Time Augmentation (TTA), contributes to improved generalization and reduced prediction variance. As demonstrated in Table II, the proposed HADE-Net outperforms baseline models in terms of both accuracy and robustness, particularly reflected in the improved Top-2 accu-racy.

These findings confirm that the combination of attention mechanisms, adaptive pooling, and ensemble learning provides a significant advantage in complex image classification tasks.

## IV. CONCLUSION

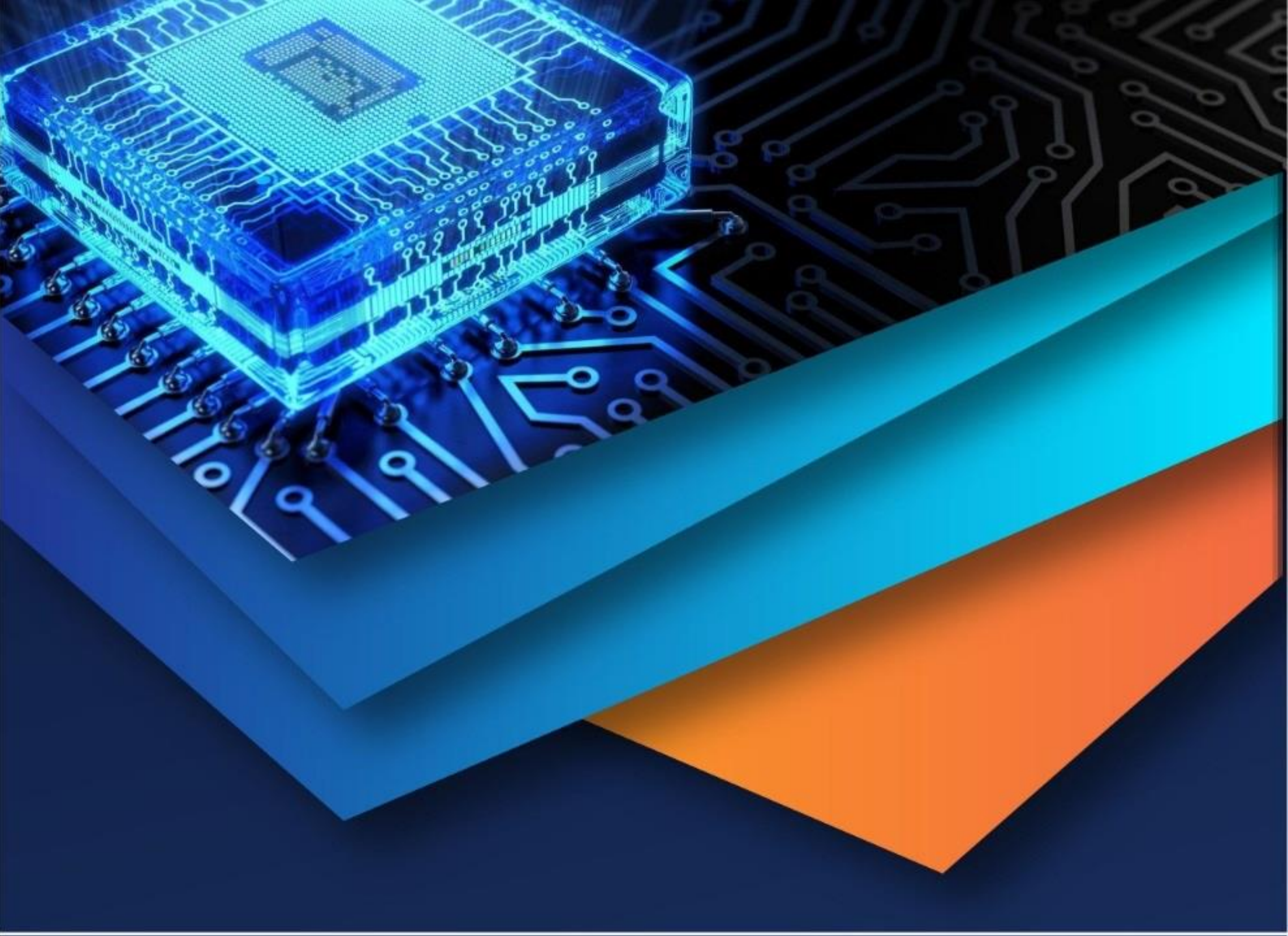
This paper presents a Hybrid Attention-Based Deep En-semble Network for flower recognition. By integrating EfficientNetV2-S and DenseNet201 with attention-based fu-sion and GeM pooling, the proposed model achieves 95.10% validation accuracy and 98.50% Top-2 accuracy. Compared to previous ensemble models, the proposed method provides improved feature representation and robustness.

### A. Future Work Includes

1) Transformer-based architectures
2) Larger datasets
3) Real-time deployment systems

## REFERENCES

[1] M. F. Rabbi et al., "An Ensemble-based Deep Learning Model for Multi-class Flower Recognition," 2023.
[2] M. Tan and Q. Le, "EfficientNet," ICML, 2019.
[3] K. He et al., "Deep Residual Learning," CVPR, 2016.
[4] G. Huang et al., "DenseNet," CVPR, 2017.
[5] D. Kingma and J. Ba, "Adam Optimizer," ICLR, 2015.
[6] N. Kumar et al., "Leafsnap: A Computer Vision System for Automatic Plant Species Identification," ECCV, 2012.
[7] M. Tan and Q. Le, "EfficientNetV2: Smaller Models and Faster Train-ing," ICML, 2021.
[8] Z.-H. Zhou, "Ensemble Methods: Foundations and Algorithms," 2012.
[9] A. Vaswani et al., "Attention Is All You Need," NeurIPS, 2017.
[10] S. Woo et al., "CBAM: Convolutional Block Attention Module," ECCV, 2018.
[11] F. Radenovic´ et al., "Fine-tuning CNN Image Retrieval with No Human Annotation," IEEE TPAMI, 2019.
[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei,"ImageNet: A Large-Scale Hierarchical Image Database," Proc. IEEE Conf. Com-puter Vision and Pattern Recognition (CVPR), 2009, pp. 248–255.
[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al.,"Going Deeper with Convolutions," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations (ICLR), 2015.
[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classifica-tion with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1097–1105.




doi crossref
10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY


Call : 08813907089 (24*7 Support on Whatsapp)