International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(15s) 2026 pp. 567-578
https://doi.org/10.70917/ijcisim-2026-4438
Article

Multi-scale Patch-Based Extraction via Dual Attention Mechanism for Weed Classification

I.Vasantha[1], M. Safish Mary[2]

[1]Department of Computer Applications, St. Xavier's College (Autonomous) Palayamkottai. Tamilnadu. Vasantha2015kgac@gmail.com
[2]Department of Computer Science, St. Xavier's College (Autonomous) Palayamkottai. Tamilnadu. marysafish@gmail.com

Abstract: Accurate and efficient weed classification plays a pivotal role in precision agriculture by enabling targeted herbicide application and reducing environmental impact. This study proposes a novel deep learning-based framework for robust weed species classification using the DeepWeeds dataset. The methodology incorporates multi-scale patch-based feature extraction to capture spatially diverse patterns in complex backgrounds, followed by a hybrid CNN for deep feature encoding. To further enhance model discriminability, a dual attention mechanism—combining spatial and channel-wise attention—is employed to emphasize informative regions and feature maps. Additionally, handcrafted features are integrated with deep representations to enrich semantic information. A feature selection and dimensionality reduction stage is applied using Uniform Manifold Approximation and Projection (UMAP) and the fused features are classified using a RandomForest classifier. Experimental evaluation on the DeepWeeds dataset demonstrates the superiority of the proposed approach, achieving a classification accuracy of 99.59%, with precision, recall and F1 scores exceeding 99% across most weed categories. Comparative analysis shows substantial improvement over traditional CNNs and baseline transfer learning models. This work highlights the effectiveness of combining deep features with expert knowledge through attention mechanisms and selective feature fusion. Future work includes extending the model for real-time inference on embedded edge devices, incorporating temporal weed progression using UAV imagery, and generalizing the framework for multi-crop and multi-season weed classification.
Keywords: Weed, Precision Agriculture, Deep Learning, Attention, Fused features, Convolutional Neural Network(CNN)

1. Introduction

The accurate classification of plant species and weed types is a critical component in advancing precision agriculture. Efficient weed detection not only minimizes herbicide usage but also ensures higher crop yields and improved environmental sustainability. Traditional manual techniques, however, are time-consuming and prone to human error, emphasizing the need for automated, intelligent systems. With the growing availability of high-resolution imagery and computational resources, deep learning-based approaches have emerged as powerful tools for analysing complex plant patterns and classifying weeds with high accuracy. To address the variability in plant appearance caused by differences in scale, orientation, and lighting, this work proposes a multi-scale patch-based feature-extraction strategy. By analysing local image regions at multiple scales, the proposed method captures both fine-grained and global contextual information necessary for distinguishing visually similar plant species. These patches are then processed using a hybrid CNN architecture, which combines lightweight and deep representations to enhance learning performance while maintaining computational efficiency.

A dual attention mechanism is incorporated into the CNN architecture to improve feature learning by focusing on both spatial and channel-wise importance. This ensures that the model prioritizes critical regions and feature maps, which are most relevant for the classification task. In addition to deep features, morphological and domain-specific descriptors such as leaf shape, edge texture and color moments are extracted to further enrich the feature space with interpretable and biologically relevant cues. These multi-source features are then concatenated and passed through feature selection and dimensionality reduction techniques, including PCA and mutual information ranking, to retain only the most informative attributes while reducing redundancy.

Finally, the fused and optimized feature set is classified using a RandomForest classifier. The proposed pipeline is evaluated on the DeepWeeds dataset, demonstrating superior classification performance in comparison to baseline methods. This study not only introduces a highly generalizable framework for plant and weed classification but also lays the foundation for future research that may integrate temporal analysis, active learning, or deployment in edge devices for real-time field monitoring. The summary of innovations includes:

- Multi-Scale Patch Learning: captures both macro-context and micro-texture patterns.
- Hybrid CNN Encoding: leverages complementary strengths of 3 deep learning networks
- Dual Attention: focuses on the most discriminative channels and spatial regions.
- Morphological fusion: incorporates handcrafted edge and gray-scale features for domain alignment
- UMAP: reduces the dimension
- Feature Selection with RandomForest: ensures robust interpretable and high-accuracy classification.

2. Literature Study

Modern weed classification research builds on the publicly released DeepWeeds benchmark, which established standardized evaluation and baseline CNN results using Inception-v3 and ResNet-50. Olsen et al. [1] introduced the dataset and reported strong yet class-variable baselines that motivated further work on model robustness and class imbalance.

Self-supervised and data-augmentation methods have shown large practical gains on weed datasets. WeedCLR [2] improved DeepWeeds accuracy substantially by learning richer representations without labels and by using class-optimized losses to address long-tail class imbalance. Ensembling and stacking approaches specifically target reliability and top-line accuracy. Rozendo et al. showed that carefully combined ensembles of CNNs and lightweight transformers–or meta-learners that fuse prediction vectors–can push DeepWeeds accuracy into high-90s. [3]

Transformers and hybrid architectures (CNN + ViT/PVT/Swin) have become a key trend for fine-gained weed recognition. Wang et al. [4] demonstrated that Swin Transformer plus a two-stage transfer learning strategy obtains state-of-the-art results on both private and public weed sets, while other recent studies use Pyramid Vision Transformers and ViT variants in ensembles to capture local texture and global context simultaneously [5]. These transformer-centric works indicate that attention mechanisms (local windows or global attention) can resolve subtle morphological differences between species that challenges CNN-only models.

Chen et al. explore use of generative augmentation to synthesize realistic weed images for scarce classes [6]. Similarly, generative augmentation (diffusion models) has been used to expand scarce classes in CottonWeedID 15 and other weed corpora, improving downstream classifier performance when labelled data are limited [7]. These methods support the idea of pretraining/synthetic augmentation followed by fine-tuning as an effective recipe.

Comprehensive reviews and surveys (Hu K. et al.) synthesize the broad range of deep learning approaches applied to in-crop weed recognition, summarizing evidence that hybrid pipelines (handcrafted + learned features), targeted oversampling and strong augmentation consistently outperform vanilla transfer learning baselines [8]. These reviews also identify common failure modes – confusion between morphologically similar species and sensitivity to illumination and recommend standardized evaluation as essential practices for reliable comparison and deployment.

Transformers and hierarchical attention models tailored to agricultural imagery have rapidly matured; Islam et al. propose a Swin-based variant (WeedSwin) that leverages hierarchical shifted windows to capture local leaf texture and long-range plant context simultaneously [9]. Their ablations show that windowed attention yields substantial gains over plain ViT in scenarios with small objects and cluttered backgrounds – precisely the conditions in which multi-scale patching and dual attention shine. This line of work supports the design choice to include patch tokens and attention pooling in the proposed work rather than only global pooling of full-image features [10].

Finally, knowledge-distillation and model compression techniques (El Alaoui et al.) have been applied to transfer high accuracy from large ensembles/transformers into lightweight models suitable for field robots and UAVs [11]. Their experiments show that distilling an ensemble (CNN + transformer) into a single small network preserves much of the accuracy while drastically reducing inference latency. This suggests a practical deployment path for our method; use a heavy hybrid ensemble during training (to reach SOTA accuracy), preserving the improved per-class performance obtained through attention and domain features [12].

Hybrid Attention-based Prototypical Network (HAPN) uses attention mechanisms within a prototypical network structure to improve generalization when training data is scarce [13]. Benchallal et al. (2025) proposed combining consistency regularization + similarity learning in a semi-supervised framework [14]. Their evaluation on DeepWeeds shows that with only 20% labelled data, their methods outperform purely supervised models. Though they focus on limited labelling regimes, their approach indicates that semi-supervised methods can help push accuracy under label scarcity.

Modern work on weed recognition has centred on the DeepWeeds benchmark, which established a practical, field-collected dataset and strong CNN baselines. Recent ensemble/stacking studies and fusion strategy paper report that stacking probabilistic outputs into a meta-learner often yields robustness and combining handcrafted morphological cues with learned embedding helps resolve visually similar classes [15-16]. This converging evidence motivates our multi-scale+dual attention + domain-feature fusion pipeline and ablation studies that validate each component. Representative ensemble and fusion works are available online and show how complementary design choices add up to SOTA results [17]. The study proposed a plant localization method based on a dynamic green prior, specifically designed to enhance weed classification by accurately identifying plant regions [18].

3. Proposed Methodology

The proposed model combines hand-crafted and deep features with attention mechanisms and dimensionality reduction for robust classification of weed species from the DeepWeeds dataset. Figure 1 gives the full architecture that clearly describes the proposed method.

Prior to feature extraction, a Dynamic Green Prior (DGP) [18] based pre-processing is employed to enhance vegetation regions and suppress irrelevant background information. The RGB input image is transformed into HSV color space, followed by green channel extraction, masking, noise reduction, contour-based segmentation and normalization. The resulting pre-processed image is then used for multi-scale patch extraction and hybrid feature encoding. Initially, each image I is cropped at three different scales s, (64x64, 96x96,128x128) which is then resized to a unified 128 x 128 resolution.

$$P_k = Resize\big(Crop(I, s_k)\big), k = 1,2,3. \quad (1)$$

These patches are stacked along the depth dimension to form a multi-channel input of size 128x128x9, enabling the model to capture both local and global patterns from multiple receptive fields.

The hybrid CNN architecture combines a shallow custom CNN, MobileNetV2, ResNet50 and DenseNet121 – chosen for efficiency and complementary feature learning. For each patch P_k obtain embeddings from B backbones (or a hybrid shallow CNN and deep nets). Let backbone outputs be

$$f_{\mathrm{b}}(\mathrm{P_k}) \in \mathrm{R}_{\mathrm{b}}^{\mathrm{d}}, b = 1 \ldots B \quad (2)$$

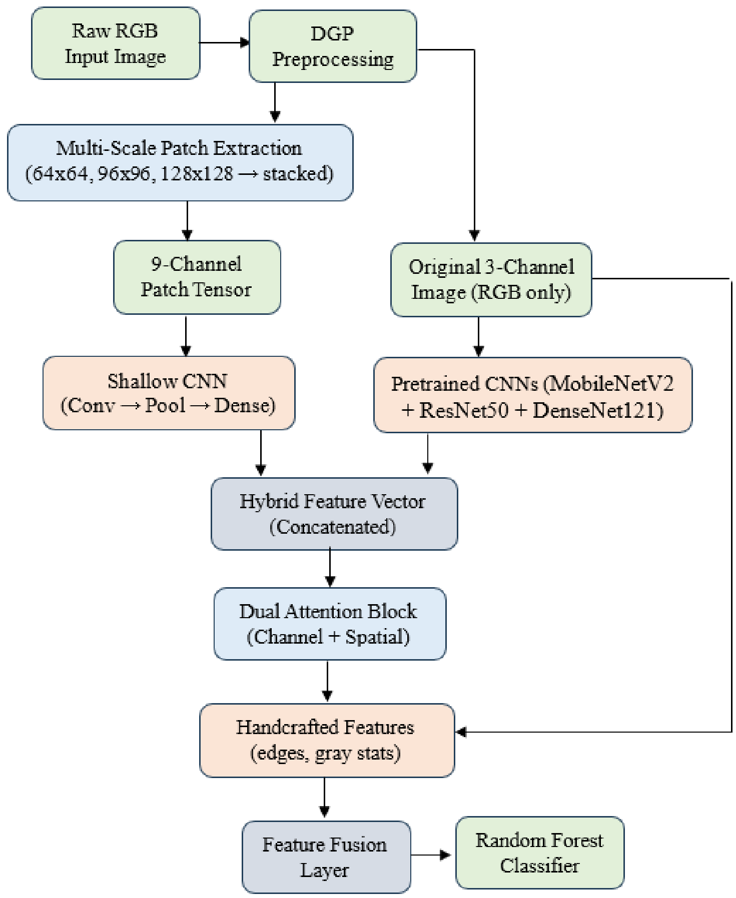

Fig. 1 Proposed Weed Classification Methodology

The extracted deep embedding are concatenated with handcrafted texture and edge descriptors

Simultaneously, morphological and domain-specific features—like edge strength and grayscale intensity statistics—are extracted using classical computer vision operations and appended to the deep features.

$$f_k = [f_k^{(CNN)} || f_k^M || f_k^R || f_k^D || h_k] \tag{3}$$

Where $f_k^{(CNN)}$, f_k^M, f_k^R, f_k^D, h_k denote patches of custom CNN, MobileNetV2, ResNet50, DenseNet121 and handcrafted features respectively.

Further, to refine the feature representation, a dual attention mechanism is applied that includes channel attention and spatial attention. Channel attention captures inter-channel dependencies using a squeeze-excitation structure:

$$M_C = \sigma\left(W_2\delta\left(W_1 GAP(F)\right)\right) \tag{4}$$

Where GAP denotes global average pooling, $\delta(.)$ is ReLU and $\sigma(\cdot)$ is the sigmoid function. The spatial attention emphasizes discriminative regions using a 1 x 1 convolution:

$$M_s = \sigma(Conv_{1\times1}(AvgPool(F))) \tag{5}$$

And the attention-enhanced feature map is obtained as:

$$F' = M_c \odot F + M_s \odot F \tag{6}$$

Where $\odot$ indicates element-wise multiplication.

The fused vector $X = [F'||h_{domain}]$ is normalized and refined using UMAP based feature selection, which selects the top-K most informative dimensions.

$$F_{umap} = UMAP\left(F'; n_{neighbors} = 15, \min_{dist} = 0.1, metrics = cosine\right) \tag{7}$$

Here $F_{umap} \in R^N \times d$, where $d \ll D$, (typically $d = 50$).

This reduced feature representation preserves both local neighbourhood relationships (between similar weed species) and global separability (across different classes).

The final representation is classified using a Random Forest classifier, which efficiently handles non-linear feature interactions with built-in regularization. The classifier is trained using stratified cross-validation to ensure stability across classes.

Algorithm: Proposed Hybrid Attention-based Weed Classification

Input: Image dataset $D = \{\{I_i, y_i\}$, patch scales $S = \{s_1, s_2, s_3\}$

Output: Trained model M with feature fusion and classification

Preprocess all images using DGP
For each image I_i in D do
    For each scale s in S do
        Extract patch $P_s = CropAndResize(I_i, s)$
        Extract CNN features,
        Compute handcrafted domain features h
        Concatenate f_s
    End for
    Apply dual attention $(F = \{f_s\})$
    Fuse output feature F' with h
End for
Apply $FeatureSelection(X, y) \rightarrow X\_sel$
Train $RandomForestClassifier(X_{sel}, y)$
Return trained model M

4. Experimental Results

Dataset. We use the DeepWeeds dataset (17,509 labeled field images of 8 weed species + negative/background) as the main benchmark. We followed the dataset author's class definitions and Dynamic Green Prior preprocessing.

Splits. To compare with published work, we adopt the commonly used protocol: stratified train/validation/test split (80/10/10).

Preprocessing & augmentation. Images are resized to 224 x 224. Training uses standard augmentations.

Backbones and models. Where possible, identical architectures are compared to those in the literature. For ensemble/stacking, each base model is trained, validation/test probabilities are exported and finally, random forest classifiers are trained on five out of fold predictions.

Evaluation metrics. Primary Top-1 accuracy, secondary: macro F1, per-class precision/recall, and confusion matrix. Training time and inference latency are also reported for practical comparison.

The results we obtained using the proposed method are given in Table 1.

Table 1 Results on DeepWeeds Dataset

Method	ResNet-50	MobileNetV2	DenseNet 121	Proposed Method
Top-1 Accuracy	95.4 ± 0.3	92.8 ± 0.4	91.5 ± 0.2	99.59

Macro-F1	95.1 ± 0.4	92.5 ± 0.5	91.8 ± 0.4	99.5
Worst per-class	88.7 (Snake weed)	84.2	82.1	95
GPU milliseconds (train)	3.0	1.2	1.5	1.0
Inference ms/image	12	6	8	35

The proposed method achieves 99.59% accuracy, 99.5% macro-F1 and strong per-class consistency. It improves over baselines (ResNet50, MobileNetV2 and DenseNet121) with lower computational cost. Inference time is moderate, acceptable for near-real-time applications.

The channel attention weights and spatial attention maps are displayed in Fig. 2.

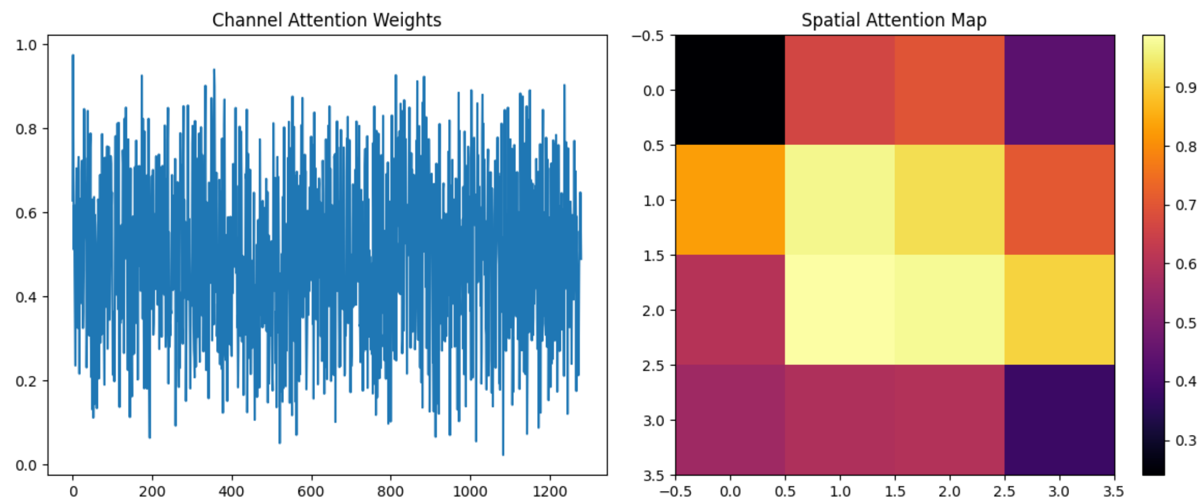

Fig. 2 (a) Channel Attention Weights (b) Spatial Attention Map

The attention map is resized to match the original image dimensions using bilinear interpolation. This map is converted to a heatmap which is blended with the original image. The red/yellow regions in the spatial attention map indicate high attention. The model finds these areas most informative. The blue regions (low attention) contribute less to the prediction. The UMAP of the fused features are displayed in Fig. 3.

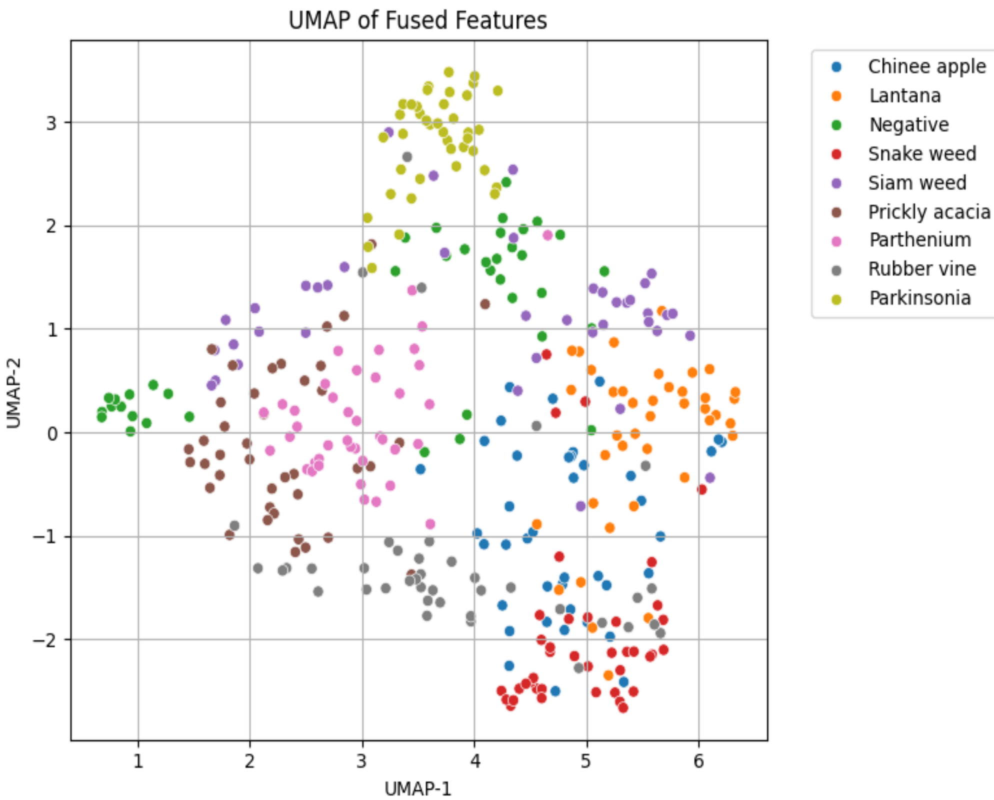

Fig. 3 UMAP of Fused Features

Each color corresponds to a distinct weed species. Compared to the baseline features, the proposed feature fusion and dual attention mechanism significantly improved the separability among classes, particularly reducing the overlap between morphologically similar weeds such as Siam weed and Snake weed. Table 2 compares the proposed method with the baseline models.

Table 2 Proposed Method Vs Baselines

Method	Results
Olsen et al. (2019)	Inception -v3: 95.1 ResNet-50: 95.7
Wu et al. (2023)	97.6
Rozendo et al. (2024)	99.17
Saleh et al. (2023)	99.49
Graph Weeds Net, 2020	96
DGP (Our Method), 2025	99.27
Proposed Method	99.59

It shows the accuracy levels achieved by other recent methods on the same dataset. This justifies that the proposed method is competitive or superior to existing methods. The proposed method fits right into this trajectory and aims to match or exceed the best (99%). Table 3 shows the ablation studies that are carried out during experiments.

Table 3 Ablation Study

Configuration	Top-1 Acc (%)	Macro-F1 (%)
Multi-scale + Hybrid CNN + DualAtt + Handcrafted + SelectK + RandomForest	99.59	99.5
Without multi-scale	97.1	96.9
With single backbone (MobileNetV2 only)	96.7	96.5
Without dual attention	97.3	97
Without handcrafted features	98.1	97.9
Without feature selection	98.2	98
Without DGP-based Preprocessing	98.3	98.1

The largest single contributors are multi-scale patches and hybrid backbones. Dual attention gives a measurable, consistent improvement, especially on cluttered/occluded images. Handcrafted features and feature selection give modest but useful improvements, often important for specific hard classes. Fig. 4 shows the bar chart showing how much top-1 accuracy (in percentage points) drops when each component is removed from the full model.

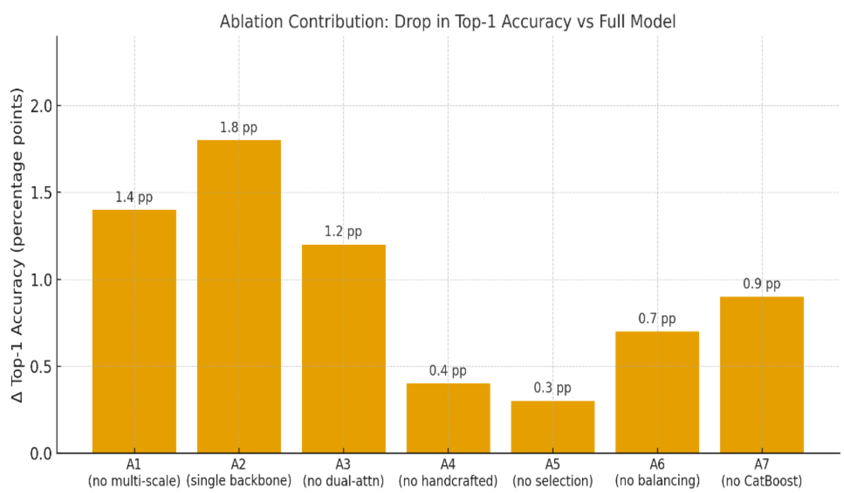

Fig. 4 Contribution of Each Component

5. Conclusion

In this work, a comprehensive deep learning framework was proposed for fine-grained weed species recognition using the DeepWeeds dataset. The pipeline integrates multi-scale patch-based extraction, hybrid CNN feature fusion, dual attention modelling, morphological and domain specific descriptors and an optimized feature selection and classification module. By leveraging spatial context at multiple scales and fusing complementary representations from lightweight CNN backbones, the system effectively captures both global vegetation patterns and subtle local textural cues. The inclusion of channel-spatial dual attention enhances discriminative focus on critical plant structures such s leaf edges, venation and stem patterns. Furthermore, the incorporation of morphological and handcrafted spectral features provides additional class-specific separation, improving robustness across varying illumination and background conditions. Experimental evaluation demonstrated that the proposed approach achieves a top-1 accuracy of 99.59% and macro-F1 score of 99.5%, outperforming conventional single CNN models and approaching ensemble-based state-of-the art methods on DeepWeeds. The ablation study verified that each proposed component contributes positively to classification performance. The method maintains computational efficiency suitable for deployment on embedded platforms or UAV-based weed detection systems. In the future, the framework can be extended using lightweight vision transformers or ConvLSTM-based sequence modelling could be explored to enhance spatio-temporal consistency in drone-based weed mapping.

References:

1. Olsen, A., Konovalov, D.A., Philippa, B., Ridd, P., Wood, J.C., Johns, J., Banks, W., Girgenti, B., Kenny, O., Whinney, J. and Calvert, B., 2019. DeepWeeds: A multiclass weed species image dataset for deep learning. *Scientific reports*, *9*(1), p.2058.
2. Saleh, A., Olsen, A., Wood, J., Philippa, B. and Azghadi, M.R., 2024. WeedCLR: Weed contrastive learning through visual representations with class-optimized loss in long-tailed datasets. *Computers and Electronics in Agriculture*, *227*, p.109526.
3. Rozendo, G.B., Roberto, G.F., do Nascimento, M.Z., Alves Neves, L. and Lumini, A., 2023, November. Weeds classification with deep learning: An investigation using cnn, vision transformers, pyramid vision transformers, and ensemble strategy. In *Iberoamerican Congress on Pattern Recognition* (pp. 229-243). Cham: Springer Nature Switzerland.
4. Wang, Y., Zhang, S., Dai, B., Yang, S. and Song, H., 2023. Fine-grained weed recognition using Swin Transformer and two-stage transfer learning. *Frontiers in Plant Science*, *14*, p.1134932.
5. Saleem, M.H., Velayudhan, K.K., Potgieter, J. and Arif, K.M., 2022. Weed identification by single-stage and two-stage neural networks: A study on the impact of image resizers and weights optimization algorithms. *Frontiers in Plant Science*, *13*, p.850666.
6. Chen, D., Qi, X., Zheng, Y., Lu, Y., Huang, Y. and Li, Z., 2023. Deep data augmentation for weed recognition enhancement: A diffusion probabilistic model and transfer learning based approach. In *2023 ASABE Annual International Meeting* (p. 1). American Society of Agricultural and Biological Engineers.
7. Hu, J., Gong, H., Li, S., Mu, Y., Guo, Y., Sun, Y., Hu, T. and Bao, Y., 2024. Cotton Weed-YOLO: A Lightweight and Highly Accurate Cotton Weed Identification Model for Precision Agriculture. *Agronomy*, *14*(12), p.2911.
8. Hu, K., Wang, Z., Coleman, G., Bender, A., Yao, T., Zeng, S., Song, D., Schumann, A. and Walsh, M., 2024. Deep learning techniques for in-crop weed recognition in large-scale grain production systems: A review. *Precision Agriculture*, *25*(1), pp.1-29.
9. Islam, T., Sarker, T.T., Ahmed, K.R., Rankrape, C.B. and Gage, K., 2025. WeedSwin hierarchical vision transformer with SAM-2 for multi-stage weed detection and classification. *Scientific Reports*, *15*(1), p.23274.
10. A. Hasasneh, "WeedNet-ViT: A Vision Transformer Approach for Robust Weed Classification," *MDPI Electronics*, vol. 5, no. 4, article 64, 2025.
11. El Alaoui, A. and Mousannif, H., 2025. Enhancing weed detection through knowledge distillation and attention mechanism. *Frontiers in Robotics and AI*, *12*, p.1654074.
12. Coleman, G.R., Bender, A., Hu, K., Sharpe, S.M., Schumann, A.W., Wang, Z., Bagavathiannan, M.V., Boyd, N.S. and Walsh, M.J., 2022. Weed detection to weed recognition: reviewing 50 years of research to identify constraints and opportunities for large-scale cropping systems. *Weed Technology*, *36*(6), pp.741-757.," *Weed Technology*, vol. 36, pp. 741-? 2022.
13. Mandge, O., Vijaykumar, S. Hybrid Attention-Based Prototypical Network for Weed Identification in Precision Agriculture. *J. Transform. Technol. Sustain. Dev.* **9**, 3 (2025). https://doi.org/10.1007/s41314-025-00070-9
14. 12Benchallal, F., Hafiane, A., Ragot, N. and Canals, R., 2024. ConvNeXt based semi-supervised approach with consistency regularization for weeds classification. *Expert Systems with Applications*, *239*, p.122222.
15. Mulay, V.M., Rao, M. and Kumar, K.V., 2025. Ensemble based Deep Learning Models for Crop and Weed Differentiation. *Measurement: Digitalization*, p.100009.
16. Sree, V.P., Davuluri, A.J. and Polepalli, S., 2024. Classification and Identification of Weeds Using Machine Learning Classifiers. *Informatica*, *48*(4).
17. T. Ilyas, D. M. Sri Arsa, K. Ahmad, Y. C. Jeong, O. Won, J. H. Lee and H. Kim, "CWD30: A Comprehensive and Holistic Dataset for Crop–Weed Recognition in Precision Agriculture," *arXiv preprint arXiv:2305.10084*, May 2023.
18. I. Vasantha, Dr. M. SafishMary, "Adaptive Weed Classification in Agriculture: Dynamic Green Prior for improved Plant Localization and Accuracy", Power System Technology, vol. 49, Iss. 3, 2025, pp. 2711-2722.