Received: 18 September 2025 | Accepted: 18 April 2026

DOI: 10.1002/agj2.70419

Agronomy Journal

ORIGINAL ARTICLE

Implementation of machine learning algorithms for accurate identification and classification of weeds

Abhinav Bhatt[1] | Shivani Kothiyal[1] | Vineeta Rathore[1] | Ankita Jha[2] | Rajeev Ranjan[1,3]

[1]G.B. Pant University of Agriculture and Technology, Pantnagar, India

[2]ICAR-Indian Institute of Water Management, Bhubaneswar, India

[3]ICAR-Central Marine Fisheries Research Institute, Kochi, India

Correspondence
Rajeev Ranjan, ICAR-Central Marine Fisheries Research Institute, Kochi, India.
Email: rajeevranjanagri@gmail.com

Assigned to Associate Editor Nothabo Dube.

[Correction added on 05 June 2026, after first online publication:Correspondence information of Rajeev Ranjan was changed from “G.B. Pant University of Agriculture and Technology, Pantnagar, India” to “ICAR-Central Marine Fisheries Research Institute, Kochi, India” and also removed the funding information.]

Abstract

The identification of weeds remains a critical challenge in modern agriculture, where invasive species compete with crops for essential resources, leading to reduced yield and quality. Advancements in computer vision and deep learning present effective prospects for automated weed detection, enhancing precision, efficiency and sustainability in agriculture. This research explores the potential of convolutional neural networks (CNNs) for weed classification, incorporating transfer learning and lightweight architectures to achieve improved efficiency. The performance of five CNN models (VGG16 [where VGG is visual geometry group], VGG19, InceptionV3, ResNet50 [where ResNet is residual networks], and MobileNet) was assessed using a dataset of *rabi* season weed species (*Coronopus didymus*, *Fumaria parviflora*, *Medicago denticulata*, and *Rumex acetosella*). Images of these weeds were captured under field conditions using a high-resolution mobile camera. The dataset comprised 14,418 images, partitioned into training, validation, and test sets. MobileNet emerged as the best-performing model, achieving 94% validation accuracy and 88.87% test accuracy, with strong precision, recall, and F1-scores across all weed classes. InceptionV3 and VGG16 also showed robust performance, while ResNet50 exhibited underfitting, suggesting a need for further optimization. The key findings emphasize MobileNet’s suitability for real-time deployment, owing to its computational efficiency and high accuracy, making it well-suited for precision herbicide applications. Future work should aim to improve model generalization under varying field conditions, addressing dataset imbalances and integrating explainable AI techniques to facilitate farmer adoption. The study contributes to advancing AI-driven solutions in agriculture, supporting the transition toward smarter and more sustainable farming practices.

Abbreviations: CI, confidence interval; CNN, convolutional neural network; CPU, central processing unit; DL, deep learning; GPU, graphic processing unit; RAM, random access memory; ResNet, residual networks; RGB, red green blue; SE, standard error; UAV, unmanned aerial vehicle; VGG, visual geometry group; YOLO, You Only Look Once.

Plain Language Summary

The current subject studies accurate weed classification and identification, which remains a challenge in current agricultural practices. The accurate and precise weed identification is possible with new technologies, thus helping in precise weed control and reduced environmental impact. The study evaluated different machine learning (ML) models through digital images to identify and understand the best ML model for accurate identification and detection of weed in *Tarai* region of Uttarakhand. The MobileNet, an ML model, was found to be the best in detecting the weed species and their occurrence areas. The study holds importance in terms of agriculture, where the farmers follow blanket application of herbicides to control weeds, but the current study can help in precise application, thus preventing the environmental damages and soil damages.

1 | INTRODUCTION

Weeds are invasive plants that compete with crops for essential resources such as nutrients, water, and sunlight, significantly reducing agricultural productivity (García-Navarrete et al., 2024). Unlike crops, which are selected and managed for yield and quality, weeds grow aggressively and adapt quickly to diverse environmental conditions, making them a persistent problem in agricultural systems (Buhler, 2002). In addition to yield reduction, weeds can adversely affect crop quality by contaminating harvested produce with seeds or vegetative parts of weeds, thereby lowering market value (Chikoye et al., 2004). Farmers incur additional costs in weed management through chemicals, mechanical control, or labor cost. These expenses can account for up to 30% of the total crop production costs in some agricultural systems (Rao, 2000). Effective weed management is crucial for sustainable farming, particularly in *rabi* (winter) season crops, where weeds like *Coronopus didymus*, *Fumaria parviflora*, *Medicago denticulata*, and *Rumex acetosella* pose major challenges (Holm et al., 1997; Rao et al., 2000). Traditional weed control methods, such as manual removal and blanket herbicide application, are labor-intensive, costly, and environmentally harmful due to excessive chemical use (Subeesh et al., 2022).

The detection of weeds remains a critical challenge in current agriculture, as these invasive species compete directly with crops for essential resources. This competition often results in reduced crop yield and deterioration of crop quality, thereby threatening food security and farmer profitability (Oerke, 2006). Traditional methods of weed classification, which rely on manual scouting, are time-consuming, labor-intensive, and prone to human error, particularly in large scale farming systems (Slaughter et al., 2008). With the increasing diversity of weed flora, coupled with the intensification of cropping systems, accurate and timely weed identification has become essential for effective management strategies. Recent advancements in artificial intelligence and machine learning provide effective tools to address these challenges, enabling early detection, precise classification, and site-specific management of different weed species, ultimately supporting sustainable and resource-efficient agriculture (dos Santos Ferreira et al., 2017; Kamilaris & Prenafeta-Boldú, 2018). Precision agriculture, powered by deep learning (DL) and computer vision, offers a promising alternative by enabling automated, targeted weed detection and management. In contrast, modern technologies enable early detection, precise classification, and site-specific weed management, providing more accurate and sustainable solutions (Slaughter et al., 2008). Computer vision techniques, driven by machine learning and DL models, play a crucial role in extracting meaningful feature extraction from images for identification and classification of weed. Convolutional neural networks (CNNs) such as You Only Look Once (YOLO), residual networks (ResNet), and InceptionV3 have been successfully applied for real-time weed detection and classification, achieving high accuracy under varied field conditions (dos Santos Ferreira et al., 2017; Kamilaris & Prenafeta-Boldú, 2018).

Recent advancements in CNNs have revolutionized weed identification in agriculture. Studies demonstrate that DL models, including YOLO, ResNet, and InceptionV3, achieve high accuracy in distinguishing weeds from crops, even in complex "green-on-green" scenarios where similarities in color and texture make differentiation difficult (Hasan et al., 2021; Venkataraju et al., 2023). For instance, Sapkota et al. (2020) used unmanned aerial vehicle (UAV)-based red green blue (RGB) imagery and machine learning to map weeds in cotton fields with over 89% accuracy, while Subeesh et al. (2022) reported 97.7% accuracy for bell pepper weed detection using InceptionV3. Hyperspectral imaging further enhances weed classification by capturing spectral signatures invisible to conventional cameras (Mensah et al., 2024). Despite these successes, significant challenges remain, particularly in addressing issues such as variability in weed morphology in different growth stages. Most studies focus on short-term, single-crop scenarios (Blank et al., 2023), and rabi-season-specific weed datasets are limited. Additionally, while transfer learning from pre-trained models improves performance (Jabir et al., 2021), real-time deployment in varying field conditions requires further optimization (García-Navarrete et al., 2024).

This research aims to develop a CNN-based (VGG16, VGG19, InceptionV3, ResNet50, and MobileNet) weed identification system tailored for *rabi* crops, leveraging existing DL architectures to address gaps in seasonal adaptability and species-specific detection. By integrating high-resolution imagery and advanced preprocessing techniques, the study seeks to enhance precision in weed localization and classification, ultimately contributing to sustainable herbicide use and higher crop yields. Substantial advancements in DL have already laid the groundwork for its use in agricultural imaging, and this is an effort to automate weed identification accurately and reliably through DL in *Tarai* region of Uttarakhand. The study emphasizes the use of high computational transfer learning, lightweight architectures, and data augmentation techniques to develop scalable and deployable weed detection systems for precision agriculture.

Core Ideas

- Different convolutional neural network (CNN) models were statistically evaluated using different measures to compare their performances on augmented and non-augmented image datasets for accurate weed identification and classification.
- Based on augmentation, MobileNet outperformed other models, achieving validation/test accuracies of 88%–97% for *Rabi* weeds (0.97/0.97 for *Rumex acetosella*).
- The MobileNet model performed well with a validation accuracy of 94%, while ResNet50 (where ResNet is residual networks) underperformed due to insufficient feature extraction.
- Comparative analysis of multiple CNN models revealed that deeper architectures such as ResNet50 benefit significantly from augmented training data.
- Though the findings showed that the augmentation helps in enhancing the model performance, it also depends on the choice of the model.
- Augmentation enhanced model performance for classes with limited samples, including *Fumaria parviflora*.

2 | MATERIALS AND METHODS

2.1 | Experimental site

The research was conducted at Norman E. Borlaug Crop Research Centre, Pantnagar, situated in the *Tarai* region at a latitudinal and longitudinal extent of 29.02˚N and 79.25˚E at an elevation of 2484 m. The site experiences annual rainfall of 1560 mm, majorly during the south-west monsoon season, with an average rainfall of 27 mm received during the *rabi* season, extending from October to March, enabling seasonal weed study in this season.

2.2 | Weed species

A systematic field survey was conducted to collect the images of prevalent weed species based on their dominance, ecological adaptability, and potential impact on crop productivity. The following weed species were identified and selected for detailed study and classification during the *rabi* season. The rabi season is an important cropping season in the *Tarai* region because during this season principal vegetable and cereal crops are grown to fulfill the food requirements and food security of the country. Thus, controlling *rabi* weeds is required to improve the yield. The selection of only specific species as given in Figure 1 of weeds has been done due to their prevalence in most of the *rabi* crops in *Tarai* region.

2.3 | Image collection for weed recognition

The digital image acquisition for weed identification was done manually (Figure 2) using Samsung Galaxy A23 5G mobile device, and the image labeling was done visually. The weed species identification was based on distinct morphological characteristics, including leaf shape, venation, margin, and growth pattern. The imaging system features a 50-megapixel primary camera sensor equipped with optical image stabilization technology. Images were captured with the camera positioned approximately 0.5–1.0 m above the plant canopy, simulating a typical handheld or low-altitude field imaging scenario, and based on the camera resolution and acquisition height, the images provided sufficient spatial detail to clearly resolve leaf texture, shape, and plant boundaries necessary for weed classification. The high-resolution sensor ensures reduced motion blur and enhanced image quality during field data collection. Images were captured primarily at near-nadir orientations, with minor deviations due to hand-

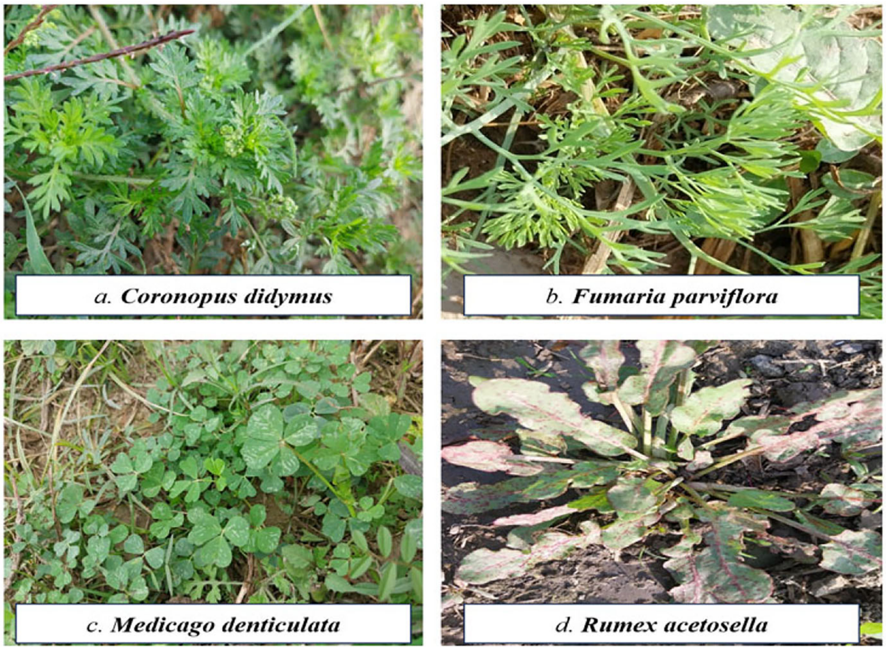

FIGURE 1 Visual records of collected weed species (*a. Coronopus didymus, b. Fumaria parviflora, c. Medicago denticulata, d. Rumex acetosella*).

FIGURE 2 *Rabi* season weed species image collection.

held acquisition, which reflects realistic field conditions while maintaining consistent visibility of plant morphology. The imaging process, as demonstrated in Figure 2, was conducted under natural daylight conditions, including morning and midday periods and both sunny and partially cloudy skies, and it provided a robust dataset suitable for training DL models for accurate weed classification and analysis, as demonstrated in Table 1 and Table 2.

2.4 | Transfer learning for weed classification

Knowledge transfer represents a methodology for disseminating expertise, techniques, and technological innovations between institutions and research domains. This approach facilitates the transformation of theoretical research into practical applications, advancing various sectors, including medical sciences, agricultural technology, and industrial manufacturing.

The test dataset was treated as an independent and unseen dataset and was not involved in any part of model training or tuning. Stratified sampling was used to maintain balanced class representation. Model selection was based on validation results, and final model performance was assessed using the test dataset only. The classification framework employed transfer learning principles combined with advanced DL architectures to create a robust weed species identification system. Multiple cutting-edge CNNs were evaluated, including VGG16, VGG19, InceptionV3, ResNet50, and MobileNet architectures. Class distribution was analyzed by computing the number and percentage of samples per weed class. The analysis revealed class imbalance, which is typical in agricultural datasets due to uneven weed occurrence in natural field conditions. The flowchart for the methodology used has been given in Figure 3.

TABLE 1 Description of rabi season weeds.

S. No.	Weed spp.	Common name	Characteristics
a)	*Coronopus didymus*	Lesser Swinecress, Bitter Cress, or Jangli Hala	• It thrives in moist, fertile soils and is highly competitive for nutrients and space (Figure 1a). • It also served as a host for pests and diseases affecting crops (Holm et al., 1997).
b)	*Fumaria parviflora*	Fineleaf Fumitory	• It prefers well-drained soils possesses medicinal properties (Rao et al., 2000) (Figure 1b).
c)	*Medicago denticulata*	Bur Clover or Maina	• It prefers well drained and fixes nitrogen, thus improving soil fertility (Figure 1c). • It becomes invasive in agricultural settings (Holm et al., 1997).
d)	*Rumex acetosella*	Red sorrel or Sheep's sorrel	• It thrives in moist, nutrient-rich soils and is used as leafy vegetables or in traditional medicine (Figure 1d). • They accumulate oxalates, which are toxic to livestock in large quantities (Radosevic et al., 2007).

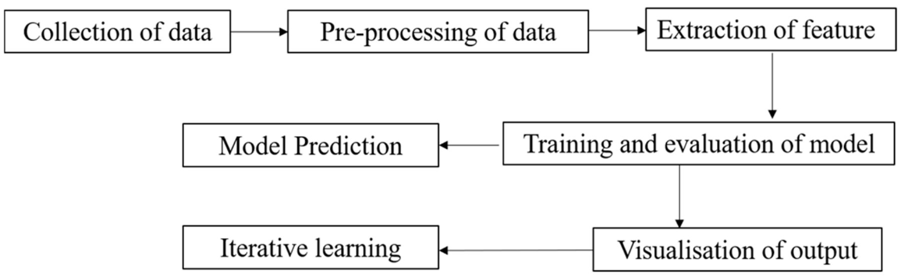

FIGURE 3 Flowchart for Artificial Intelligence work in Deep Learning Models.

TABLE 2 Number of images taken for each weed.

Weeds	No. of images taken
Coronopus didymus	3508
Fumaria parviflora	3516
Medicago denticulata	3799
Rumex acetosella	3601

2.5 | Different models used for training and evaluation

2.5.1 | VGG16 and VGG19

The VGG16 model has been developed by Oxford University's Visual Geometry Group in 2014 with deep CNN having straightforward and consistent design and architecture, demonstrating effective performance in image classification tasks (Mascarenhas & Agarwal, 2021). Similarly, the VGG19 has been built upon the VGG16 foundation with enhanced depth, and it has incorporated layers of 19 weighted, 16 convolutional, and 3 fully connected layers (Mascarenhas & Agarwal, 2021).

2.5.2 | InceptionV3

The InceptionV3 model was created by Google's research team in 2015, representing an advanced CNN that enhances the original Inception (GoogLeNet) framework through architectural refinements (Xia et al., 2017).

2.5.3 | ResNet50

The ResNet50 model uses revolutionary skip connections (residual connections), unlike VGG architectures that enable training of significantly deeper networks. This Microsoft Research innovation, introduced in 2015, addresses the vanishing gradient problem in deep neural networks (Mascarenhas & Agarwal, 2021).

2.5.4 | MobileNet

The MobileNet model has been developed by Google in 2017, which addresses the computational constraints of mobile and embedded systems through efficient architectural design. This lightweight model prioritizes reduced computational requirements, memory usage, and power consumption while maintaining acceptable accuracy levels (Sinha & El-Sharkawy, 2019).

2.6 | Training and validation of models

All experiments were implemented using TensorFlow (version 2.x) with the Keras high-level API, which provides a stable and widely adopted framework for CNN training and evaluation. Data augmentation was implemented using the Keras image preprocessing utilities, applied on-the-fly during training. Model training was performed on a system equipped with an NVIDIA graphic processing unit (GPU) (with CUDA support), an Intel/Advanced Micro Device multi-core CPU, and at least 16 GB RAM. Training time per model ranged from 30 to 90 min, depending on network depth. Inference time per image was in the order of milliseconds, enabling near-real-time performance. The Adam optimizer was used due to its adaptive learning rate capability and fast convergence, which is particularly suitable for training deep CNNs on moderately sized datasets. An initial learning rate of 0.0001 was used. Early stopping was not applied; models were trained for a fixed number of epochs. Dropout layers were included in the classification head to reduce overfitting. No explicit weight decay was applied. 10 epochs were used, and the maximum number of epochs were kept consistent across models to ensure fair comparison. The training of the models was done using processed datasets in batches of 32 images, with standardization of each image to a dimension of 224 × 224 pixels. The *rabi* season weed classification task was performed on the following dataset:

Number of training images used: 10,089

Number of validation images used: 2883

Number of testing images used: 1446

The dataset were split into training, validation, and test sets using a random sampling strategy, ensuring that the class distribution of weed species was preserved across all subsets. This approach prevents bias toward dominant classes and allows fair performance evaluation across minority classes. Dataset splitting was performed before data augmentation. Augmentation was applied only to the training set to increase data diversity while ensuring that validation and test sets remained representative of real, unseen field conditions. To improve model generalization and robustness to field variability, several geometric and photometric augmentation techniques were applied to the training data. These included random rotations, horizontal and vertical flips, brightness and contrast adjustment, zooming, and random cropping. Augmentation parameters were selected to reflect realistic field variations and included random rotations within ± 20°, horizontal and vertical flips, brightness and contrast adjustments within ± 20%, zooming up to 10%, and random cropping while preserving the central plant region. Data augmentation was applied exclusively to the training set. Validation and test datasets were kept unchanged to ensure unbiased performance evaluation on real, unseen images. Image augmentation was performed on-the-fly during training, meaning that each epoch presented a different augmented version of the same image. As a result, no fixed number of augmented images per original image was stored. To assess the impact of data augmentation, models were trained and evaluated using both augmented and non-augmented training datasets, while keeping the validation and test sets identical. Performance metrics demonstrated consistent improvement when augmentation was applied.

6.2 The confusion matrix provides a comprehensive n × n tabular representation of model performance, with rows indicating actual classifications and columns representing predicted classifications (Beauxis-Aussalet & Hardman, 2014). Model evaluation metrics, including accuracy, precision, recall, and F1-score, were computed using scikit-learn. Confusion matrices were also generated using scikit-learn utilities. An error analysis was conducted using confusion matrices and per-class performance metrics, which revealed that most misclassifications occurred between morphologically similar weed species. In addition, confidence intervals for the model accuracy metrics were used to demonstrate the additional robustness or lack of the models.

Actual values/actual values	Predicted positive	Predicted negative
Actual positive	True positive (TP)	False negative (FN)
Actual negative	False positive (FP)	True negative (TN)

Mathematically standard classification performance metrics such as Accuracy, Precision, Recall, F1-Score, Confidence Intervals, Standard Error (SE) can be expressed as

2.6.1 | Accuracy

Accuracy measures the ratio of correct predictions to total predictions, providing a fundamental performance indicator.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

2.6.2 | Precision

Precision evaluates the reliability of positive predictions by calculating the fraction of correctly identified positive instances among all positive prediction.

$$\text{Precision}_{\text{class A}} = \frac{T\,P_{\text{class A}}}{T\,P_{\text{class A}} + F\,P_{\text{class A}}}$$

2.6.3 | Recall

Recall (sensitivity) measures the model's ability to identify all actual positive instances within the dataset.

$$\text{Recall}_{\text{class A}} = \frac{T\,P_{\text{class A}}}{T\,P_{\text{class A}} + F\,N_{\text{class A}}}$$

2.6.4 | F1 score

The F1-score provides a balanced metric combining precision and recall through harmonic mean.

$$\text{F1}_{\text{class A}} = 2X\frac{\text{Precision}_{\text{class A}}\ X\ \text{Recall}_{\text{class A}}}{\text{Precision}_{\text{class A}} + \text{Recall}_{\text{class A}}}$$

2.6.5 | Standard error

$$\text{SE} = \text{SQRT}\left(\frac{\text{Overall accuracy} \times (1 - \text{overall accuracy})}{\text{Total number of predictions}}\right)$$

2.6.6 | Confidence interval (85% accuracy)

$$\text{CI} = \text{Overall accuracy} \pm 1.96 \times \text{SE}$$

3 | RESULTS

Each model's performance was rigorously analyzed using multiple evaluation criteria such as accuracy, precision, recall, and F1-score. After extensive experimentation and comparison, the model that excelled across all these metrics was chosen as the top performer. Model accuracy indicates the overall correctness of predictions; precision measures the proportion of correctly identified positive samples out of all predicted positives; recall reflects the model's ability to capture actual images; and F1-score provides a balanced assessment by harmonizing precision and recall. The strengths and weaknesses of each model were systematically assessed across different weed species and classification scenarios using metrics in combination. The top DL-performing model was selected based on this multi-criteria evaluation framework, thereby providing the most reliable solution for weed classification under real-world agricultural conditions.

3.1 | MobileNet

The MobileNet model shows strong training performance, with training accuracy increasing steadily and reaching nearly 98% by the 10th epoch. However, the validation accuracy fluctuates between 92% and 94%, showing a clear gap between training and validation performance. This suggests that the model is overfitting; it learns well on training data but does not generalize as well on unseen data. Training loss decreases consistently and reaches a very low value (~0.05), indicating the model is fitting the training data very well. In contrast, the validation loss does not show a steady decline and even increases after certain epochs, further supporting the overfitting observations illustrated in Figure 4.

3.1.1 | Validation and test classification report

The validation performance of the MobileNet model, as shown in Table 3, was assessed using precision, recall, and F1-score, and it demonstrated consistently strong results across all weed classes. *Coronopus didymus* achieved an F1-score of 0.92 with high recall (0.97) and precision of 0.88, indicating effective detection with minimal missed cases. *Fumaria parviflora* showed balanced performance with an F1-score of 0.91, precision of 0.95, and recall of 0.87, while *M. denticulata* recorded an F1-score of 0.94, supported by high precision (0.95) and recall (0.92). The best performance was observed for *R. acetosella*, which achieved an F1-score of 0.97, reflecting near-perfect recall (0.99) and high precision (0.96).

The test results as presented in Table 3 further confirmed the robustness of the MobileNet model, with an overall F1-score of 0.89. Across classes, the model maintained a good balance between precision, recall, and F1-scores. *Coronopus didymus* exhibited high recall (0.96) with a moderate precision (0.77), resulting in an F1-score of 0.85. *Fumaria parviflora* recorded a precision of 0.89 and recall of 0.78, yielding an F1-score of 0.83. For *M. denticulata*, the model achieved excellent precision (0.98) with a recall of 0.83, leading to an F1-score of 0.90. Consistent with validation results, *R. acetosella* again showed outstanding performance, with precision of 0.95, recall of 0.98, and an F1-score of 0.97. The macro and weighted average scores (≈0.89–0.90) indicate stable and reliable classification across weed categories.

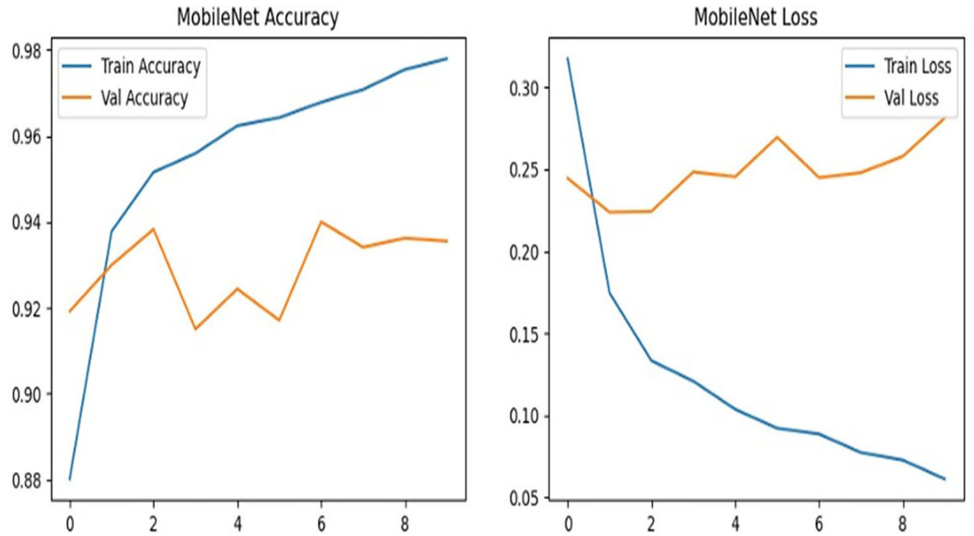

FIGURE 4 Training and validation accuracy and loss curves of MobileNet for different epochs.

TABLE 3 Classification report of validation in MobileNet.

Weed class	Precision (validation)	Recall (validation)	F1-score (validation)	Precision (test)	Recall (test)	F1-score (test)
Coronopus didymus	0.88	0.97	0.92	0.77	0.96	0.85
Fumaria parviflora	0.95	0.87	0.91	0.89	0.78	0.83
Medicago denticulata	0.95	0.92	0.94	0.98	0.83	0.9
Rumex acetosella	0.96	0.99	0.97	0.95	0.98	0.97
Overall accuracy	0.935	0.9375	0.935	0.8975	0.8875	0.8875

Overall, MobileNet achieved validation and test accuracies of 94% and 89%, respectively, demonstrating strong robustness, generalization capability, and suitability for practical weed species identification in precision agriculture.

3.1.2 | Validation performance (confusion matrix)

To further assess the classification performance, the confusion matrix for the validation dataset was examined. The matrix indicates that MobileNet achieved strong accuracy across all four weed classes as presented in Table 4. For example, *C. didymus* was correctly classified 678 times out of a total of 701 samples, with only a few misclassifications distributed among other classes. *Fumaria parviflora* also showed robust performance, achieving 609 correct predictions, despite some overlap being observed with closely related classes. *Medicago denticulata* was accurately classified 697 times, though there were minor misclassifications mainly with *F. parviflora* and *R. acetosella*. Finally, *R. acetosella* exhibited the highest precision, with 713 correct predictions out of 720 samples, reflecting the model's strong capability (99.03%) to distinguish this species. These outcomes confirm the effectiveness of MobileNet in weed species classification.

3.1.3 | Test performance (confusion matrix)

The confusion matrix for the MobileNet model demonstrates strong classification performance across the four weed species. *Coronopus didymus* was accurately predicted in 339 out of 352 cases, with only a few misclassifications, primarily as *F. parviflora* (13 cases), as presented in Table 5. *Fumaria parviflora* also showed robust accuracy, with 273 correct predictions and limited confusion, largely with *C. didymus* and occasional errors involving *M. denticulata* and *R. acetosella*. *Medicago denticulata* had a slightly higher misclassification rate, with 318 correct identifications, with some samples misclassified as *C. didymus*, *F. parviflora*, and *R. acetosella*. *Rumex acetosella* achieved excellent prediction accuracy, correctly classified in 355 out of 361 samples, with very few misclassifications. Overall, most errors were minor and generally involved species with similar features, reflecting the model's high reliability in distinguishing these weed species.

TABLE 4 Validation confusion matrix of MobileNet.

Actual/predicted	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**678** **96.58%** **True**	19 2.71% False	2 0.29% False	2 0.29% False
Fumaria parviflora	65 9.24% False	**609** **86.62%** **True**	28 3.98% False	1 0.14% False
Medicago denticulata	23 3.03% False	12 1.58% False	**697** **91.82%** **True**	27 3.56% False
Rumex acetosella	3 0.42% False	1 0.14% False	3 0.42% False	**713** **99.03%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

TABLE 5 Confusion matrix of MobileNet on test dataset.

Actual/predicted	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**339** **95.49%** **True**	13 3.66% False	0 00.00% False	0 00.00% False
Fumaria parviflora	77 21.91% False	**273** **77.56%** **True**	1 0.28% False	1 0.28% False
Medicago denticulata	27 7.18% False	18 4.79% False	**318** **84.57%** **True**	18 4.79% False
Rumex acetosella	0 00.00% False	2 0.55% False	4 1.11% False	**355** **98.34%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

The misclassifications primarily indicate areas where further refinement might improve feature differentiation, especially involving *M. denticulata*as.

3.2 | InceptionV3

The InceptionV3 model demonstrated noticeable improvement in both training and validation accuracy. The training accuracy increased to approximately 94%, while the validation accuracy peaked around 91%, as illustrated in Figure 5. Although a gap between training and validation accuracy is evident, it is smaller than that observed for MobileNet, indicating moderate overfitting. The training loss consistently decreased across epochs, whereas the validation loss fluctuated around 0.35–0.40 without substantial improvement after the initial epochs. This behavior suggests certain generalization limitations, though overall learning trends remain stable.

3.2.1 | Validation and test classification report

The classification performance of the InceptionV3 model was evaluated on both validation and test datasets using precision, recall, and F1-score metrics, as summarized in Table 6. On the validation dataset, the model demonstrated consistently strong performance across all weed species. *Coronopus didymus* achieved an F1-score of 0.88, supported by a high recall of 0.93 and a precision of 0.84, indicating reliable detection with few missed samples. *Fumaria parviflora* exhibited a balanced performance, attaining an F1-score of 0.88 with precision and recall values of 0.92 and 0.84, respectively. Similarly, *M. denticulata* recorded an F1-score of 0.88, reflecting strong classification capability with precision of 0.91 and recall of 0.85. The best validation performance was observed for *R. acetosella*, which achieved an F1-score of 0.93, supported by precision of 0.91 and recall of 0.95. Overall, the InceptionV3 model achieved a validation accuracy of approximately 89%,

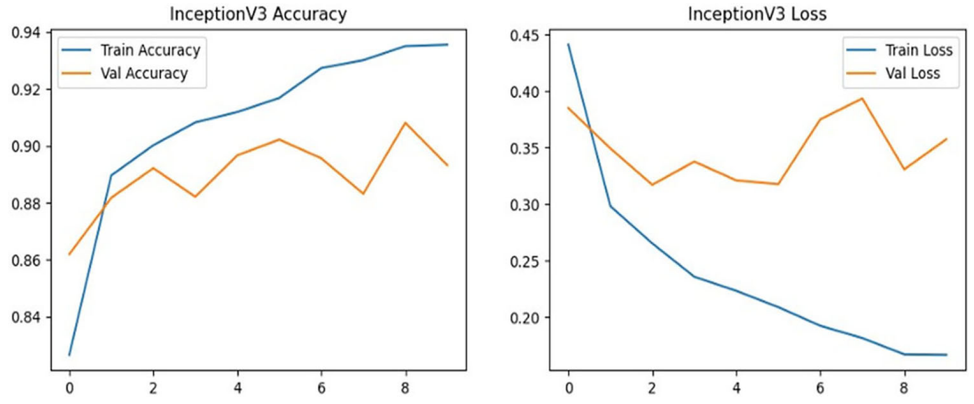

FIGURE 5 Training and validation accuracy and loss curves of InceptionV3 for different epochs.

TABLE 6 Classification report of validation in InceptionV3Weed class.

	Precision (validation)	Recall (validation)	F1-score (validation)	Precision (test)	Recall (test)	F1-score (test)
Coronopus didymus	0.84	0.93	0.88	0.66	0.97	0.79
Fumaria parviflora	0.92	0.84	0.88	0.84	0.67	0.74
Medicago denticulata	0.91	0.85	0.88	0.97	0.7	0.81
Rumex acetosella	0.91	0.95	0.93	0.94	0.98	0.96
Overall accuracy	0.895	0.8925	0.8925	0.8525	0.83	0.825

with macro and weighted average F1-scores close to 0.89, demonstrating balanced classification performance across all weed categories.

On the test dataset, the InceptionV3 model maintained stable performance, achieving an overall accuracy of approximately 82.6% and an F1-score of 0.83. *Coronopus didymus* exhibited high recall (0.97) but lower precision (0.66), resulting in an F1-score of 0.79, indicating effective detection with some false positives. *Fumaria parviflora* achieved an F1-score of 0.74, with moderate recall (0.67) and relatively high precision (0.84). *Medicago denticulata* showed excellent precision of 0.97 but a recall of 0.70, yielding an F1-score of 0.81. Consistent with validation results, *R. acetosella* demonstrated outstanding performance on the test set, achieving an F1-score of 0.96 with precision of 0.94 and recall of 0.98. The macro and weighted average F1-scores (approximately 0.82–0.83) indicate reliable and consistent classification performance across classes.

3.2.2 | Validation performance (confusion matrix)

The confusion matrix for the validation dataset, presented in Table 7, further illustrates the classification capability of the InceptionV3 model. *Coronopus didymus* was correctly classified in 655 out of 701 samples, with minimal misclassification into other classes. *Fumaria parviflora* achieved 593 correct predictions, although some confusion with other weed species was observed. *Medicago denticulata* recorded 645 correct classifications, with a small proportion of misclassifications (7.37%), primarily with *R. acetosella*. *Rumex acetosella* demonstrated the highest validation accuracy, with 682 correct predictions (94.46%) and minimal confusion. Overall, the validation confusion matrix confirms the model's robust ability to distinguish between weed species.

3.2.3 | Test performance (confusion matrix)

The test confusion matrix for InceptionV3, shown in Table 8, reflects strong classification performance across all four weed species. *Coronopus didymus* was correctly identified in 341 out of 352 samples, with minor confusion primarily with *F. parviflora*. *Fumaria parviflora* achieved 236 correct predictions, with most misclassifications occurring as *C. didymus*. *Medicago denticulata* recorded 266 correct predictions but exhibited some confusion with *C. didymus*, *F. parviflora*, and *R. acetosella*. *Rumex acetosella* achieved the highest test accuracy, with 352 correct predictions and minimal misclassification. These results indicate that most errors occurred between species with visually similar characteristics.

TABLE 7 Validation confusion matrix of InceptionV3.

Actual/predicted	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**655** **93.05%** **True**	42 5.97% False	2 0.28% False	2 0.28% False
Fumaria parviflora	72 10.25% False	**593** **84.33%** **True**	31 4.41% False	7 1.00% False
Medicago denticulata	50 6.58% False	8 1.05% False	**645** **84.87%** **True**	56 7.37% False
Rumex acetosella	7 0.97% False	3 0.42% False	28 3.88% False	**682** **94.46%** **True**
Weed class	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

TABLE 8 Confusion matrix of InceptionV3 on test dataset.

Actual/predicted	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**341** **96.33%** **True**	11 3.11% False	0 00.00% False	0 00.00% False
Fumaria parviflora	109 30.97% False	**236** **67.05%** **True**	3 0.85% False	4 1.14% False
Medicago denticulata	65 17.06% False	31 8.14% False	**266** **69.82%** **True**	19 4.99% False
Rumex acetosella	0 00.00% False	4 1.11% False	5 1.39% False	**352** **97.51%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

3.3 | VGG16

The VGG16 model exhibited a strong and parallel increase in both training and validation accuracy. Training accuracy reached approximately 90%, while validation accuracy closely followed at around 88%, as illustrated in Figure 6. The narrow gap between the curves suggests good generalization and minimal overfitting. Both training and validation losses decreased steadily, although the validation loss declined at a slightly slower rate, indicating stable learning behavior.

3.3.1 | Validation and test classification report

The classification performance of the VGG16 model on both validation and test datasets is summarized in Table 9. On the validation dataset, the model demonstrated strong and balanced performance across all weed species. *Coronopus didymus* achieved an F1-score of 0.89, supported by precision of 0.93 and recall of 0.84. *Fumaria parviflora* recorded an F1-score of 0.87, with higher recall (0.92) indicating strong sensitivity. *Medicago denticulata* also achieved an F1-score of 0.87, reflecting balanced precision (0.89) and recall (0.85). *Rumex acetosella* showed the best validation performance, with an F1-score of 0.92 and nearly equal precision and recall values. Overall, the VGG16 model achieved a validation accuracy of approximately 89%, with macro and weighted average F1-scores close to 0.89.

On the test dataset, VGG16 achieved an accuracy of approximately 78.5% and an overall F1-score of 0.78. *Coronopus didymus* achieved an F1-score of 0.72, primarily affected by lower recall. *Fumaria parviflora* demonstrated strong recall (0.88) and an F1-score of 0.75, while *M. denticulata* achieved an F1-score of 0.78, supported by good precision (0.86). *Rumex acetosella* again achieved the highest test per-

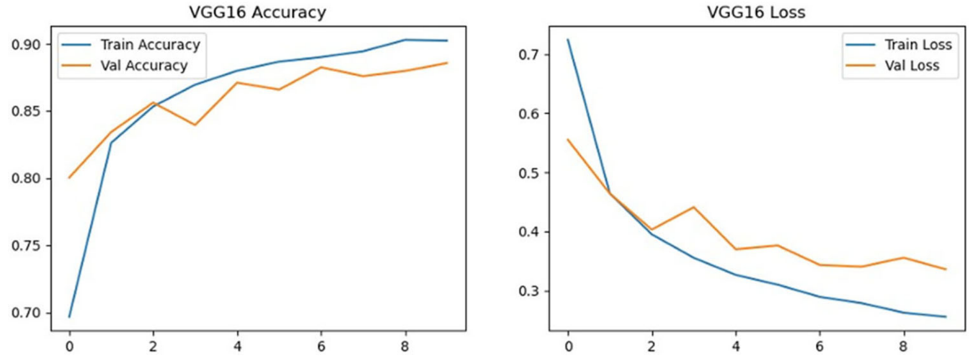

FIGURE 6 Training and validation accuracy and loss curves of VGG16 (where VGG is visual geometry group) for different epochs.

TABLE 9 Classification report of validation in VGG16 (where VGG is visual geometry group).

Weed class	Precision (validation)	Recall (validation)	F1-score (validation)	Precision (test)	Recall (test)	F1-score (test)
Coronopus didymus	0.93	0.84	0.89	0.79	0.66	0.72
Fumaria parviflora	0.82	0.92	0.87	0.65	0.88	0.75
Medicago denticulata	0.89	0.85	0.87	0.86	0.71	0.78
Rumex acetosella	0.92	0.93	0.92	0.9	0.89	0.89
Overall accuracy	0.89	0.885	0.8875	0.8	0.785	0.785

TABLE 10 Validation confusion matrix of VGG16 (where VGG is visual geometry group).

Acutal/predicted	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**592** **84.33%** **True**	98 13.96% False	11 1.57% False	0 00.00% False
Fumaria parviflora	18 2.56% False	**650** **92.59%** **True**	30 4.27% False	5 0.71% False
Medicago denticulata	20 2.65% False	41 5.43% False	**642** **84.92%** **True**	56 7.41% False
Rumex acetosella	4 0.56% False	7 0.97% False	40 5.56% False	**669** **92.91%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

formance, with an F1-score of 0.89. The macro and weighted average scores confirm consistent but comparatively lower test performance than validation results.

3.3.2 | Validation performance (confusion matrix)

The validation confusion matrix for the VGG16 model, presented in Table 10, indicates strong classification accuracy across all weed species. *Coronopus didymus* was correctly classified in 592 samples, with limited misclassification. *Fumaria parviflora* achieved 650 correct predictions, showing minimal confusion. *M. denticulata* recorded 642 correct classifications, though some confusion with *R. acetosella* was observed. *Rumex acetosella* achieved 669 correct predictions, demonstrating strong validation performance. Overall, the matrix confirms the model's effectiveness during validation.

3.3.3 | Test performance (confusion matrix)

The test confusion matrix for VGG16, shown in Table 11, reveals increased interclass confusion compared to validation results. *Coronopus didymus* was correctly identified in 234 samples, with notable misclassification as *F. parviflora*.

TABLE 11 Confusion matrix of VGG16 (where VGG is visual geometry group) on test dataset.

Actual/predicted	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**234** **66.10%** **True**	107 30.23% False	7 1.98% False	4 1.13% False
Fumaria parviflora	37 10.51% False	**310** **88.07%** **True**	5 1.42% False	0 00.00% False
Medicago denticulata	25 6.56% False	53 13.91% False	**271** **71.13%** **True**	32 8.40% False
Rumex acetosella	0 00.00% False	8 8.40% False	33 8.40% False	**320** **88.40%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

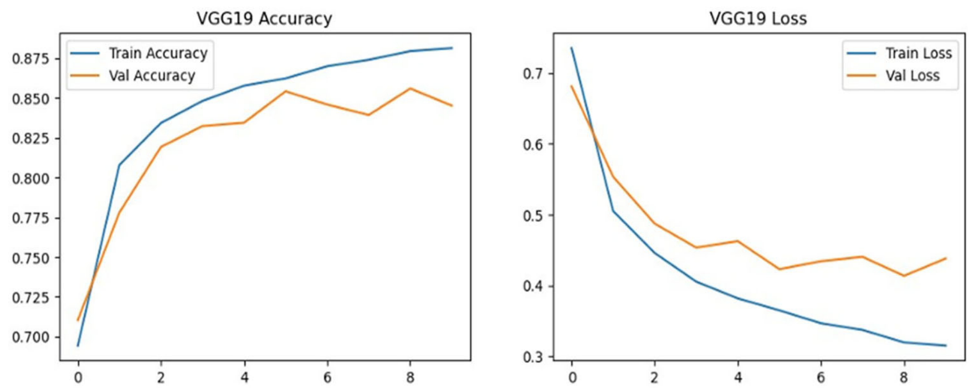

FIGURE 7 Training and validation accuracy and loss curves of VGG19 (where VGG is visual geometry group) for different epochs.

Fumaria parviflora achieved 310 correct predictions, while *M. denticulata* showed moderate confusion with other classes. *Rumex acetosella* achieved the highest test accuracy with 320 correct classifications. These results indicate that while VGG16 performs well, further optimization could reduce interclass confusion.

3.4 | VGG19

The VGG19 model showed consistent improvement throughout training. Training accuracy reached approximately 88%, while validation accuracy peaked near 85%, as illustrated in Figure 7. Both training and validation losses decreased gradually with a small and stable gap, indicating balanced learning and minimal overfitting.

3.4.1 | Validation and test classification report

The classification performance of VGG19 on validation and test datasets is presented in Table 12. On the validation dataset, the model demonstrated balanced performance across all weed species. *Coronopus didymus* achieved an F1-score of 0.85, with high precision and slightly lower recall. *Fumaria parviflora* recorded an F1-score of 0.84, supported by strong recall. *Medicago denticulata* achieved an F1-score of 0.81, indicating moderate sensitivity. *Rumex acetosella* achieved the highest validation F1-score of 0.88, supported by high recall. Overall, the model achieved a validation accuracy of approximately 85% with a macro F1-score of 0.84.

On the test dataset, VGG19 achieved an overall accuracy of approximately 76% and an F1-score of 0.75. *Rumex acetosella* again achieved the best test performance with an F1-score of 0.87. *Fumaria parviflora* achieved an F1-score of 0.73, while *C. didymus* and *M. denticulata* showed lower recall values, resulting in moderate F1-scores. The macro and weighted averages indicate moderate and uneven performance across classes.

3.4.2 | Validation performance (confusion matrix)

The validation confusion matrix for VGG19, shown in Table 13, indicates strong classification performance, partic-

TABLE 12 Classification report of validation in VGG19 (where VGG is visual geometry group).

Weed class	Precision (validation)	Recall (validation)	F1-score (validation)	Precision (test)	Recall (test)	F1-score (test)
Coronopus didymus	0.9	0.8	0.85	0.76	0.62	0.69
Fumaria parviflora	0.79	0.9	0.84	0.63	0.88	0.73
Medicago denticulata	0.92	0.73	0.81	0.94	0.6	0.73
Rumex acetosella	0.81	0.96	0.88	0.81	0.94	0.87
Overall accuracy	0.855	0.8475	0.845	0.785	0.76	0.755

TABLE 13 Validation confusion matrix of VGG19 (where VGG is visual geometry group).

Actual/ predicted	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**558** **79.60%** **True**	138 19.69% False	2 0.29% False	3 0.43% False
Fumaria parviflora	27 3.79% False	**632** **90.17%** **True**	33 4.64% False	11 1.54% False
Medicago denticulata	29 3.82% False	27 3.55% False	**553** **72.76%** **True**	150 19.74% False
Rumex acetosella	3 0.42% False	7 0.97% False	16 2.22% False	**694** **96.39%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

TABLE 14 Confusion matrix of VGG19 (where VGG is visual geometry group) on test dataset.

Weed class	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**220** **62.50%** **True**	129 36.65% False	1 0.28% False	2 0.57% False
Fumaria parviflora	38 10.80% False	**309** **87.78%** **True**	2 0.57% False	3 0.85% False
Medicago denticulata	30 7.89% False	45 11.84% False	**229** **60.26%** **True**	77 20.26% False
Rumex acetosella	0 00.00% False	9 2.49% False	11 3.05% False	**341** **94.46%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

ularly for *R. acetosella* and Fumaria parviflora. *Coronopus didymus* and *M. denticulata* exhibited higher misclassification rates, especially between each other. Overall, the validation matrix confirms reliable performance with some class overlap.

3.4.3 | Test performance (confusion matrix)

The test confusion matrix for VGG19, presented in Table 14, demonstrates that *R. acetosella* achieved the highest accuracy, with minimal confusion. *Fumaria parviflora* also showed

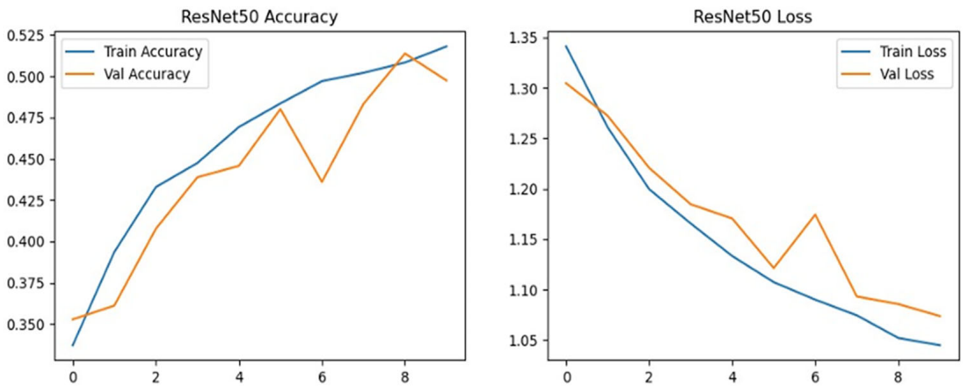

FIGURE 8 Training and validation accuracy and loss curves of ResNet50 (where ResNet is residual networks) for different epochs.

TABLE 15 Classification report of validation in ResNet50 (where ResNet is residual networks).

Weed class	Precision (validation)	Recall (validation)	F1-score (validation)	Precision (test)	Recall (test)	F1-score (test)
Coronopus didymus	0.49	0.58	0.53	0.38	0.43	0.4
Fumaria parviflora	0.45	0.24	0.31	0.41	0.47	0.44
Medicago denticulata	0.4	0.46	0.43	0.53	0.46	0.5
Rumex acetosella	0.63	0.71	0.67	0.71	0.63	0.67
Overall accuracy	0.4925	0.4975	0.485	0.5075	0.4975	0.5025

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

strong performance, while *C. didymus* and *M. denticulata* exhibited notable interclass confusion. These findings suggest that while VGG19 performs well for certain species, further refinement is needed for consistent classification across all classes.

3.5 | ResNet50

The ResNet50 model exhibited significantly lower performance compared to the other architectures. Training accuracy reached only about 52%, while validation accuracy remained volatile and consistently lower, as shown in Figure 8. Both training and validation losses decreased gradually but remained high, indicating underfitting and limited learning capacity with the current configuration.

3.5.1 | Validation and test classification report

The classification performance of ResNet50 on validation and test datasets is summarized in Table 15. On the validation dataset, the model showed moderate and inconsistent performance across weed species. *Rumex acetosella* achieved the highest validation F1-score of 0.67, followed by *C. didymus* with an F1-score of 0.53. *Medicago denticulata* and *F. parviflora* exhibited lower F1-scores, particularly *F. parviflora* due to very low recall. Overall, the model achieved a validation accuracy of approximately 50% and a macro F1-score of 0.48.

On the test dataset, ResNet50 achieved an accuracy of approximately 49.7% and an overall F1-score of 0.50. *Rumex acetosella* again achieved the best performance with an F1-score of 0.67, while *M. denticulata* recorded an F1-score of 0.50. *C. didymus* and *F. parviflora* showed weaker performance, resulting in lower reliability across classes.

3.5.2 | Validation performance (confusion matrix)

The validation confusion matrix for ResNet50, presented in Table 16, indicates substantial interclass confusion. *Coronopus didymus* and *F. parviflora* were frequently misclassified as each other, while *M. denticulata* showed confusion with *R. acetosella*. *Rumex acetosella* achieved comparatively better validation accuracy, though misclassification remained present.

TABLE 16 Validation confusion matrix of ResNet50 (where ResNet is residual networks).

Weed class	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**408** **58.29%** **True**	160 22.86% False	128 18.29% False	5 0.71% False
Fumaria parviflora	282 40.11% False	**166** **23.61%** **True**	218 31.02% False	37 5.26% False
Medicago denticulata	113 14.89% False	43 5.67% False	**346** **45.59%** **True**	257 33.86% False
Rumex acetosella	38 5.28% False	3 0.42% False	165 22.92% False	**514** **71.39%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

TABLE 17 Confusion matrix of ResNet50 (where ResNet is residual networks) on test dataset.

Weed class	*Coronopus didymus*	*Fumaria parviflora*	*Medicago denticulata*	*Rumex acetosella*
Coronopus didymus	**150** **42.61%** **True**	182 51.85% False	19 5.40% False	1 0.28% False
Fumaria parviflora	142 40.34% False	**165** **46.88%** **True**	43 12.22% False	2 0.57% False
Medicago denticulata	70 18.37% False	46 12.07% False	**177** **46.46%** **True**	88 23.10% False
Rumex acetosella	32 8.89% False	10 2.78% False	92 25.56% False	**227** **63.06%** **True**

Note: The bold values represents the true classification, number of correct classifications, and accuracy percentage of correct classifications performed by considered deep learning models.

3.5.3 | Test performance (confusion matrix)

The test confusion matrix for ResNet50, shown in Table 17, further confirms limited classification performance. High levels of confusion were observed between *C. didymus* and *F. parviflora*, as well as between *M. denticulata* and *R. acetosella*. These results indicate that ResNet50, in its current configuration, is not optimal for weed species classification without further tuning or dataset enhancement.

3.5.4 | Statistical evaluation of CNN models on validation and test dataset

The different CNNs were evaluated using overall classification accuracy, complemented by statistical uncertainty and significance measures as shown in Table 18. In addition to point estimates of accuracy, SEs and 95% confidence intervals were computed to assess the robustness of each CNN model. Among the five CNN models, MobileNet achieved the highest overall accuracy (93% and 89%) along with the lowest SE (0.004 and 0.008) and the narrowest 95% confidence interval for both the validation and test dataset, respectively, indicating superior predictive performance and greater stability. The corresponding p-value ($p < 0.001$) demonstrates that this model's accuracy is highly unlikely to be due to random chance, confirming its statistical robustness. InceptionV3, VGG16, and VGG19, respectively, also exhibited strong performance, with accuracies (validation: 89,89 & 84%; Test: 83,78 & 76%) close to that of MobileNet but with slightly wider confidence intervals (ranging 87%–90% for validation and 74%–84% for test dataset). All three models remained statistically significant ($p < 0.001$), indicating consistent and reliable classification performance across the dataset. In contrast, ResNet50 showed comparatively lower accuracy (50% for both datasets) and a wider confidence interval, suggest-

TABLE 18 Model accuracy metrics for validation and test dataset.

CNN Models	Total number of predictions	Overall accuracy (%)	SE	95% CI	*p*-value
Validation dataset					
MobileNet	2883	93.55	0.004	93–94	<0.01
InceptionV3	2883	89.32	0.006	88–90	<0.01
VGG16	2883	88.55	0.006	87–90	<0.01
VGG19	2883	84.53	0.007	83–86	<0.01
ResNet50	2883	49.74	0.009	48–51	<0.01
Test dataset					
MobileNet	1446	88.87	0.008	87–90	<0.01
InceptionV3	1446	82.64	0.01	81–84	<0.01
VGG16	1446	78.49	0.011	76–81	<0.01
VGG19	1446	76.00	0.011	74–78	<0.01
ResNet50	1446	49.72	0.013	47–52	<0.01

Abbreviations: CI, confidence interval; CNN, convolutional neural network; ResNet, residual networks; SE, standard error; VGG, visual geometry group.

ing higher variability in predictions. Nevertheless, its *p*-value still indicated statistically significant performance relative to the baseline classifier, confirming that even the least accurate CNN model outperformed random classification. Overall, the comparative analysis demonstrates that deeper or more optimized CNN architectures provide not only higher accuracy but also greater robustness, as reflected by narrower confidence intervals. The narrow confidence interval reflects high precision due to the large sample size. Given the large sample size of weed, statistical significance is complemented with confidence intervals and weed class-wise performance metrics to emphasize practical significance.

These results highlight the importance of evaluating CNN models using both performance metrics and statistical validation to ensure reliable and robust conclusions. All CNNs achieved statistically significant accuracies ($p < 0.01$); however, MobileNet consistently outperformed the other CNNs, exhibiting higher accuracy and lower uncertainty.

4 | DISCUSSION

To improve detection sensitivity and accuracy, Chen et al. (2024) incorporated attention mechanisms and hybrid loss functions into CNN architectures. Similarly, Srinivasu et al. (2021) focused on architectural modifications to enhance feature learning capabilities in DL models. These innovations are particularly beneficial for weed identification in *rabi* season crops, where subtle visual distinctions can hinder classification accuracy. W. Wang et al. (2020) and Rybczak and Kozakiewicz (2024) emphasized the importance of data augmentation and transfer learning in boosting the performance of lightweight CNN models. Their studies highlighted how these techniques improve model generalization, especially under conditions of limited training data common in *rabi* crop scenarios. The current study explored different CNN models (VGG16, VGG19, InceptionV3, ResNet50, and MobileNet) for weed classification. The models were tested for their accuracy in weed identification by incorporating architectures to improve their efficiency in detecting *rabi* season weed species (*C. didymus*, *F. parviflora*, *M. denticulata*, and *R. acetosella*) in *Tarai* region of Uttarakhand. The CNNs performance was also validated using accuracy and 95% confidence intervals, and the results were found to be statistically significant ($p < 0.001$), thus determining the robustness of the CNNs. The above CNN architectures were selected to represent different design approaches. VGG16 and VGG19 are deep networks that learn features in a step-by-step manner. InceptionV3 is designed to extract features at multiple scales, while ResNet50 uses residual connections to improve learning in deeper networks. MobileNet is a lightweight model that uses depth-wise separable convolutions and requires less computational power. Thus, using this variety of models helps to study how differences in network complexity, feature learning, and computational efficiency influence weed classification accuracy. This is especially important for field images, where weeds and crops appear under varying light conditions, complex backgrounds, and similar plant shapes. The CNN architectures observed confusion between *C. didymus* and *F. parviflora*, which can be attributed to their similar early growth-stage morphology, particularly in leaf size, color, and fine-textured appearance under field conditions. Both species exhibit small, dissected leaves and low-growing growth habits, which reduce distinctive visual cues in RGB imagery. Murad et al. (2023) also noted dataset imbalance as a critical limitation in CNN performance.

Goceri (2021) and Karar et al. (2021) underscored the efficiency of MobileNet in managing complex visual classification tasks, showcasing its suitability for agricultural conditions that vary across seasons. Their work demonstrated how lightweight CNNs can maintain high performance even in resource-constrained environments. The weed detection in the current study was possible with the MobileNet architecture that achieved strong accuracy with 94% validation and 89% test accuracy, demonstrating its robustness, reliability,, and effective generalization for weed species classification under diverse agricultural conditions.

The InceptionV3 model demonstrated strong and reliable classification performance across all classes in the validation set. Furthermore, Sowmiya et al. (2024) leveraged a hybrid approach combining InceptionV3 and VGG16, reaching a testing accuracy of 98.87%. This highlights the potential of integrating multiple architectures to capture diverse image features. Subeesh et al. (2022) identified InceptionV3 as the most accurate model for weed detection in bell pepper fields, with a precision rate of 97.7%. Further to advance this, Han et al. (2022) and Maeda-Gutiérrez et al. (2025) applied regularization and explainability techniques, such as Shapley Additive Explanation, to enhance model interpretability and performance, an essential step toward real-world deployment in agricultural settings. Yang et al. (2023) developed a modified VGG16 model incorporating SE attention mechanisms for weed detection in corn fields, illustrating how tailored CNNs can enhance feature discrimination in complex plant environments. All the studies reported high accuracy, underscoring VGG16's versatility and reliability across diverse applications in the field of agriculture. The InceptionV3 and VGG16 models showed robust performance in r*abi* weed species detection in the study area with an overall F1-score of 0.83 and 0.78, respectively. Avuçlu (2023) expanded their classification research to include r*abi* crops, enhancing the seasonal applicability of their approach. Al Sahili and Awad (2022) evaluated the AgriNet-pretrained VGG19 model across various crop cycles, confirming its robustness and adaptability. Attri et al. (2023) emphasized the need for up to date DL techniques in smart farming and recommended regularization and transfer learning to optimize models like VGG19 for field deployment. Y. Wang et al. (2024) optimized VGG16 for targeted weed segmentation, aiding precision spraying efforts. Y. Wang et al. (2024) optimized the VGG16 architecture for targeted weed segmentation, enhancing the model's ability to distinguish weed patches from crop regions with higher spatial accuracy. The VGG19 model in the study area performed well in *rabi* weed detection with minimal overfitting.

The ResNet50 model demonstrated considerably lower performance than the other architectures, achieving an overall average accuracy of approximately 50%. Alirezazadeh et al. (2024) compare several CNNs, including ResNet50, MobileNet, and mobile device vision transformer. The ResNet variants perform decently, but the lightweight networks outperform them overall, especially on small, imbalanced weed data sets. Similarly, ResNet50 was outperformed by MobileNetV3-Small and VGG16 in some test settings, particularly with smaller or more varied image backgrounds (Roseno et al., 2024). ResNet50 has consistently shown promise in agricultural applications. Senthilkumar and Kamarasan (2021) explored hybrid CNN architectures to enhance classification precision in agricultural tasks. Murad et al. (2023), in their review of 61 studies, found that YOLO, VGG, and ResNet were the most frequently applied models for real-time weed detection. The ResNet50 in the present study too showed underfitting, representing its meager contribution to accurate Rabi weed detection in the study area. Thus, the study emphasized the suitability and accuracy of MobileNet in real-time deployment for precision herbicide applications. Lightweight DL models such as MobileNet are well suited for agricultural image analysis, especially when the available training data are limited or contain noise. Because MobileNet has a smaller number of parameters, it reduces the chances of overfitting, which is a common problem in field-based agricultural datasets. MobileNet uses depthwise separable convolutions, which help the model learn important spatial and feature-based information more efficiently. This design allows the model to perform reliably under varying field conditions, such as changes in lighting, partial crop or weed coverage, and complex backgrounds commonly observed in agricultural environments. The classified results can later be combined with more advanced modules such as weed localization or field-level mapping. Because MobileNet has fewer parameters and lower energy requirements, it can run on edge devices like embedded GPU platforms or modern smartphones. This reduces dependency on high-end computing systems and makes the approach practical for field-level agricultural applications where power availability, cost, and response time are important. The high classification accuracy achieved in this study indicates that false spraying can be minimized. This is important for improving both economic efficiency and environmental safety in precision spraying systems. Although this study does not directly measure herbicide savings or economic benefits, the results clearly show that the proposed weed classification models have strong potential for integration into real-time precision spraying systems used in modern agriculture.

4.1 | Limitations

The study is limited by the relatively small number of major weed species present in the selected locality. As a result, the collection and labeling of field images involved some practical challenges. The machine learning algorithm cannot fully represent real field conditions, including different growth stages, background variations, and changing environmental factors. The proposed models are expected to generalize to weed species with similar morphological characteristics and imaging conditions; however, direct generalization to unseen species cannot be guaranteed without additional training data.

4.2 | Future work

This study shows good performance under the given experimental conditions, but the dataset used has some limitations. The images were collected from a single geographic region during a limited time period and include only a fixed set of weed species. Weed appearance changes across seasons, growth stages, and locations due to differences in climate, soil type, rainfall, and agricultural practices. These factors are important in meteorology and agriculture, as environmental conditions strongly affect crop and weed growth. Using images from only one region or season may reduce the model's ability to perform well in other areas or under different weather conditions.

In future work, this limitation can be addressed by collecting weed images from multiple seasons, different growth stages, and various agroclimatic regions. This can be done by conducting field surveys across different locations or by combining local data with publicly available weed image datasets. Such dataset expansion will help the model learn more diverse visual patterns. Including multi-season and multi-region data in future work is expected to reduce dataset bias, improve generalization, and make the model more reliable for real-world precision agriculture applications under varying meteorological conditions.

5 | CONCLUSION

The study evaluated multiple CNN models for weed classification, comparing their performance on both augmented and non-augmented datasets. MobileNet emerged as the top performer, achieving 94% validation accuracy, but showed signs of overfitting, highlighting the need for augmentation to improve generalization. InceptionV3 and VGG variants demonstrated balanced performance, with VGG16 showing the least overfitting, suggesting robustness in handling non-augmented data. ResNet50 underperformed, likely due to insufficient feature extraction, emphasizing the importance of augmentation for deeper models. Augmentation techniques notably enhanced model accuracy for classes with limited training samples, such as *F. parviflora*, reducing misclassification. However, over-reliance on augmented data risks masking inherent dataset biases, necessitating a balanced approach. The findings underscore that while augmentation techniques effectively enhance model performance, the ultimate success of weed classification tasks depends largely on the choice of the underlying model. For instance, optimal model selection like MobileNet stands out as a lightweight architecture suitable for resource-constrained environments such as mobile devices or field-based applications, where rapid and efficient deployment is essential. On the other hand, VGG16 provides greater stability and consistent accuracy, making it more reliable for real-world agricultural conditions or controlled agricultural monitoring scenarios, albeit with higher computational requirements. This study demonstrates that the chosen model is both practical and scalable for precision agriculture applications. This study supports precision spraying efforts by enabling site-specific herbicide application, thereby reducing chemical usage, minimizing environmental impact, and lowering production costs while maintaining effective weed control.

AUTHOR CONTRIBUTIONS

Abhinav Bhatt: Data curation; formal analysis; investigation; resources; software; validation; visualization; writing—original draft. **Shivani Kothiyal**: Formal analysis; investigation; supervision; validation; visualization; writing—original draft. **Vineeta Rathore**: Data curation; formal analysis; methodology; supervision; writing—review and editing. **Ankita Jha**: Data curation; formal analysis; resources; visualization; writing—review and editing. **Rajeev Ranjan**: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; supervision; validation; writing—review and editing.

ACKNOWLEDGMENTS

The authors are highly thankful to GBPUA&T, Pantnagar, and ICAR-CMFRI for providing all necessary facilities and support for this study. Research was supported by the Indian Council of Agricultural Research, Department of Agricultural Research and Education, Government of India.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The datasets are available on reasonable request.

ORCID

Shivani Kothiyal https://orcid.org/0000-0002-8748-2035
Rajeev Ranjan https://orcid.org/0000-0002-4856-058X

REFERENCES

Alirezazadeh, P., Schirrmann, M., & Stolzenburg, F. A. (2024). A comparative analysis of deep learning methods for weed classification of high-resolution UAV images. *Journal of Plant Diseases and Protection*, *131*, 227–236. https://doi.org/10.1007/s41348-023-00814-9

Al Sahili, Z., & Awad, M. (2022). The power of transfer learning in agricultural applications: AgriNet. *Frontiers in Plant Science*, *13*(1), 92–107.

Attri, I., Awasthi, L. K., Sharma, T. P., & Rathee, P. (2023). A review of deep learning techniques used in agriculture. *Ecological Informatics*, *77*(1), 102117.

Avuçlu, E. (2023). Classification of pistachio images using VGG16 and VGG19 deep learning models. *International Scientific and Vocational Studies Journal*, *7*(2), 79–86.

Beauxis-Aussalet, E., & Hardman, L. (2014). Simplifying the visualization of confusion matrix. *BNAIC*, *6*(1), 148.

Blank, L., Rozenberg, G., & Gafni, R. (2023). Spatial and temporal aspects of weeds distribution within agricultural fields—A review. *Crop Protection*, *17*(2), 106–200.

Buhler, D. D. (2002). Challenges and opportunities for integrated weed management. *Weed Science*, *50*(3), 273–280. https://doi.org/10.1614/0043-1745(2002)050[0273:CAAOFI]2.0.CO;2

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chen, H., Zhou, G., He, W., Duan, X., & Jiang, H. (2024). Classification and identification of agricultural products based on improved MobileNetV2. *Scientific Reports*, *14*(1), 3454.

Chikoye, D., Manyong, V. M., Carsky, R. J., Ekeleme, F., Gbehounou, G., & Ahanchede, A. (2004). Response of speargrass (*Imperata cylindrica*) to cover crops integrated with hand-weeding and chemical control in maize and cassava. *Crop Protection*, *21*(2), 145–156. https://doi.org/10.1016/S0261-2194(01)00115-1

Dos Santos Ferreira, A., Matte Freitas, D., Gonçalves Da Silva, G., Pistori, H., & Theophilo Folhes, M. (2017). Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, *143*, 314–324. https://doi.org/10.1016/j.compag.2017.10.027

García-Navarrete, O. L., Correa-Guimaraes, A., & Navas-Gracia, L. M. (2024). Application of convolutional neural networks in weed detection and identification: A systematic review. *Agriculture*, *14*(4), 568.

Goceri, E. (2021). Diagnosis of skin diseases in the era of deep learning and mobile technology. *Computers in Biology and Medicine*, *13*(1), 104–138.

Han, S., Liu, J., Zhou, G., Jin, Y., Zhang, M., & Xu, S. (2022). InceptionV3-LSTM: A deep learning net for the intelligent prediction of rapeseed harvest time. *Agronomy*, *12*(2), 30–46.

Hasan, A. M., Sohel, F., Diepeveen, D., Laga, H., & Jones, M. G. (2021). A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, *18*(4), 106–167.

Holm, L., Doll, J., Holm, E., Pancho, J., & Herberger, J. (1997). *World weeds: Natural histories and distribution*. Wiley.

Jabir, B., Falih, N., Sarih, A., & Tannouche, A. (2021). A strategic analytics using convolutional neural networks for weed identification in sugar beet fields. *AGRIS On-Line Papers in Economics and Informatics*, *13*(1), 49–57.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Karar, M. E., Alsunaydi, F., Albusaymi, S., & Alotaibi, S. (2021). A new mobile application of agricultural pests recognition using deep learning in cloud computing system. *Alexandria Engineering Journal*, *60*(5), 4423–4432.

Maeda-Gutiérrez, V., Oropeza-Valdez, J. J., Reveles-Gómez, L. C., Padron-Manrique, C., Resendis-Antonio, O., Solís-Sánchez, L. O., Guerrero-Osuna, H. A., & Olvera Olvera, C. A. (2025). AI-driven plant health assessment: A comparative analysis of inception V3, ResNet50 and ViT with SHAP for accurate disease identification in taro. *Agronomy*, *15*(1), 77.

Mascarenhas, S., & Agarwal, M. (2021). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification. *ICRTMDR*, *3*(1), 96–99.

Mensah, B., Rai, N., Betitame, K., & Sun, X. (2024). Advances in weed identification using hyperspectral imaging: A comprehensive review of platform sensors and deep learning techniques. *Journal of Agriculture and Food Research*, *18*(1), 101–108.

Murad, N. Y., Mahmood, T., Forkan, A. R. M., Morshed, A., Jayaraman, P. P., & Siddiqui, M. S. (2023). Weed detection using deep learning: A systematic literature review. *Sensors*, *23*(7), 3670.

Oerke, E. C. (2006). Crop losses to pests. *Journal of Agricultural Science*, *144*(1), 31–43. https://doi.org/10.1017/S0021859605005708

Radosevic, S. R., Holt, J. S., & Ghersa, C. M. (2007). *Ecology of weeds and invasive plants*. Wiley.

Rao, V. S. (2000). *Principles of weed science*. Oxford & IBH Publishing.

Rao, V. S., Thomasson, J. A., Hanks, J., & Wooten, J. (2000). Principles of weed science. *Computers and Electronics in Agriculture*, *60*(1), 31–38.

Roseno, M., Oktarina, S., Nearti, Y., Syaputra, H., & Jayanti, N. (2024). Comparing CNN models for rice disease detection: ResNet50, VGG16, and MobileNetV3-small. *Journal of Information Systems and Informatics*, *6*(3), 2099–2109. https://doi.org/10.51519/journalisi.v6i3.86

Rybczak, M., & Kozakiewicz, K. (2024). Deep machine learning of MobileNet, efficient and inception models. *Algorithms*, *17*(3), 96.

Sapkota, B., Singh, V., Cope, D., Valasek, J., & Bagavathiannan, M. (2020). Mapping and estimating weeds in cotton using unmanned aerial systems-borne imagery. *AgriEngineering*, *2*(2), 350–366.

Senthilkumar, C., & Kamarasan, M. (2021). An effective citrus disease detection and classification using deep learning based Inception ResNet V2 model. *Turkish Journal of Computer and Mathematics Education*, *12*(2), 283–296.

Sinha, D., & El-Sharkawy, M. (2019). Thin MobileNet: An Enhanced MobileNet Architecture. In *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 0280–0285). IEEE. https://doi.org/10.1109/uemcon47517.2019.8993089

Slaughter, D. C., Giles, D. K., & Downey, D. (2008). Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture*, *61*(1), 63–78. https://doi.org/10.1016/j.compag.2007.05.008

Sowmiya, B., Saminathan, K., & Devi, M. C. (2024). An ensemble of transfer learning based InceptionV3 and VGG16 models for paddy leaf disease classification. *Transactions on Computer and Information Technology*, *18*(1), 89–100.

Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*, *21*(8), 2852.

Subeesh, A., Bhole, S., Singh, K., Chandel, N. S., Rajwade, Y. A., Rao, K. V. R., Kumar, S. P., & Jat, D. (2022). Deep convolutional neural network models for weed detection in polyhouse grown bell peppers. *Artificial Intelligence in Agriculture*, *6*, 47–54.

Venkataraju, A., Arumugam, D., Stepan, C., Kiran, R., & Peters, T. (2023). A review of machine learning techniques for identifying weeds in corn. *Smart Agricultural Technology*, *3*, 100102.

Wang, Y., Li, Z., & Zhang, H. (2024). Optimized VGG16 for targeted weed segmentation in precision agriculture. *Journal of Agricultural Robotics*, *5*(1), 45–58.

Wang, W., Li, Y., Zou, T., Wang, X., You, J., & Luo, Y. (2020). A novel image classification approach via dense-MobileNet models. *Mobile Information Systems*, *20*(1), 160–184.

Xia, X., Xu, C., & Nan, B. (2017). Inception-v3 for flower classification. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)* (pp. 783–787). IEEE.

Yang, L., Xu, S., Yu, X., Long, H., Zhang, H., & Zhu, Y. (2023). A new model based on improved VGG16 for corn weed identification. *Frontiers in Plant Science*, *14*(1), 120–151.

How to cite this article: Bhatt, A., Kothiyal, S., Rathore, V., Jha, A., & Ranjan, R. (2026). Implementation of machine learning algorithms for accurate identification and classification of weeds. *Agronomy Journal*, *118*, e70419. https://doi.org/10.1002/agj2.70419