

Enhanced weed and crop species classification using optimized machine learning and ensemble techniques

Sathya R.*, Thirunavukkarasu K.S.

Department of Computer
Science, Vels Institute of Science,
Technology, and Advanced Studies
(VISTAS), Chennai 6200117, India

*E-mail: sathya.r0714@gmail.com

Received for publication
February 01, 2025.

Abstract

Accurate identification of weed and crop species is essential for implementing precision agriculture practices that minimize chemical usage and improve yield. This study presents an advanced machine learning framework that employs optimized classification techniques to distinguish between multiple plant species based on image features. Utilizing a dataset comprising 11,500 images from diverse agricultural classes—including crops such as maize and sugar beet, and weeds like thistle and wild oat—the framework extracts color, texture, and shape features for high-resolution pattern recognition. Several supervised machine learning algorithms, including eXtreme gradient boosting, light gradient boosting machine, and ensemble stacking methods, are explored and fine-tuned using feature selection techniques such as recursive feature elimination (RFE) and principal component analysis (PCA). The experimental results demonstrate superior performance, with the best models achieving classification accuracy of up to 96.1% and reduced inference times suitable for real-time agricultural applications. These findings highlight the potential of optimized machine learning pipelines in advancing sustainable weed management and automated field monitoring.

Keywords

machine learning, ensemble techniques, agriculture, recursive feature elimination, principal component analysis

1. Introduction

Precision agriculture has emerged as a transformative approach in modern farming, enabling efficient resource utilization, real-time monitoring, and data-driven decision-making to improve productivity and sustainability. One of the fundamental challenges in precision agriculture is the accurate identification and classification of plant species, particularly distinguishing between crop plants and invasive weeds. Weeds compete with crops for nutrients, light, and space, often resulting in significant yield losses [1–4]. Effective weed management is crucial to maintaining agricultural output, and it depends heavily on timely

and accurate detection. Traditionally, farmers have relied on manual observation and broad-spectrum herbicide application, which not only increases labor costs and chemical usage but also leads to environmental degradation and herbicide resistance. In this context, developing intelligent, automated systems for weed and crop classification can significantly enhance the efficiency and sustainability of agricultural practices.

The advent of computer vision and machine learning has opened new avenues for automating plant species recognition. Machine learning algorithms can be trained to identify specific patterns in image data, such as color, shape, texture, and spatial features,

enabling them to differentiate between species with high accuracy [5–9]. Unlike manual techniques, these methods can process vast amounts of data in real-time, making them highly suitable for deployment in autonomous systems like drones, robots, and smart spraying machines. However, the effectiveness of such systems depends on the robustness and generalization capabilities of the underlying models. Variations in lighting, occlusion, soil background, and plant growth stages present additional challenges in real-world applications, necessitating the use of advanced algorithms that can adapt to complex field environments [10–14].

This study addresses these challenges by developing a comprehensive machine learning framework that leverages optimized classification techniques and ensemble learning to enhance the recognition accuracy of weed and crop species. The dataset used in this study comprises 11,500 high-resolution images, collected from two major sources: the publicly available Plant Seedlings Classification dataset on Kaggle and a custom-curated weeds dataset. It includes 10 crop species such as maize, sugar beet, and chickweed, as well as 5 weed types including thistle, crabgrass, and wild oat. Each image is labeled with the corresponding plant species and was captured under natural lighting conditions to simulate real-world scenarios. The diversity of the dataset ensures that the developed models can generalize well across different environments and plant appearances [15–22].

The proposed framework emphasizes the use of feature-rich image representations and robust machine learning algorithms. Initially, raw images are preprocessed through resizing, normalization, and background filtering to remove noise and enhance feature clarity. Hand-crafted features such as color histograms, local binary patterns (LBP), Haralick texture descriptors, and shape-based metrics are extracted to capture the unique characteristics of each plant species. These features are essential for classical machine learning algorithms, which rely on meaningful data representations rather than end-to-end learning. To further improve the efficiency of the model, feature selection techniques are employed to reduce dimensionality while retaining the most informative attributes. Recursive feature elimination (RFE) and principal component analysis (PCA) are used in combination to select optimal subsets of features that contribute most to classification accuracy.

A variety of supervised machine learning algorithms are evaluated in this study. Decision trees and support vector machines (SVM) are initially considered due to their interpretability and effectiveness

in classification tasks. However, their limitations in scalability and performance under complex feature interactions motivate the use of more advanced models such as gradient boosting machines (GBM), eXtreme Gradient Boosting (XGBoost), and light gradient boosting machine (LightGBM). These ensemble methods combine multiple weak learners to form a strong predictive model, offering better generalization and higher accuracy. Additionally, ensemble stacking is explored by combining the outputs of multiple base classifiers through a meta-model, enhancing the model's ability to capture diverse decision boundaries. The major contributions of this research are as follows:

A comprehensive machine learning framework is designed that integrates handcrafted features (color, texture, and shape), dimensionality reduction techniques, and ensemble classifiers to accurately differentiate crop and weed species.

The study applies and compares three prominent feature selection approaches—PCA, mutual information (MI), and RFE—to reduce redundancy and enhance classification performance, especially for minority weed classes.

A novel ensemble architecture based on stacking is implemented, combining random forest (RF), SVM, XGBoost, and LightGBM. This ensemble surpasses individual model performance in both accuracy and generalization.

The model is evaluated on a large-scale, heterogeneous dataset comprising 11,500 annotated images collected from public and custom sources, representing 14 species under real-world conditions.

The framework ensures fast inference (under 80 ms/image) and uses lightweight, interpretable algorithms, making it suitable for integration into drones, embedded systems, or edge computing platforms in precision farming.

II. Literature Review

Automated classification of weeds and crops has become a critical research focus in precision agriculture due to its potential to reduce manual labor, optimize herbicide usage, and increase overall crop productivity. Recent developments in image analysis, machine learning, and remote sensing have led to increasingly sophisticated systems capable

of recognizing plant species under diverse agricultural conditions. Studies have explored both classical machine learning techniques and modern deep learning architectures, each offering unique advantages and challenges depending on the data, computational resources, and deployment environment.

Deep learning, especially convolutional neural networks (CNNs), has demonstrated exceptional performance in plant image classification tasks. Zhuang et al. [1] evaluated various deep convolutional neural networks (DCNNs) for detecting broadleaf weed seedlings in wheat. Their study highlighted the superior accuracy of DCNN models but also emphasized the challenges of computational complexity and the need for large labeled datasets. Similarly, Picon et al. [2] proposed a deep learning-based semantic segmentation approach for detecting multiple weed and corn species. Their framework utilized both synthetic and real datasets to enhance model generalization, showing that deep learning models could perform well in complex field environments when adequately trained.

Complementing image-based learning, multimodal data fusion techniques have also been introduced. Gai et al. [3] explored the fusion of color and depth images to improve crop plant detection for robotic weed control. Their findings indicate that combining spatial and spectral data enhances object separation, especially when occlusion and overlapping leaves are present. While deep learning dominates many of these approaches, studies like this demonstrate the potential for integrating handcrafted and learned features to improve model robustness.

Despite the promise of deep learning, traditional machine learning remains highly relevant, particularly when computational efficiency and interpretability are essential. Al-Badri et al. [4] provided an extensive review of machine learning techniques for weed classification, discussing the strengths and limitations of models such as SVM, RF, k-nearest neighbors (k-NN), and decision trees. The authors noted that while deep learning models tend to outperform in accuracy, classical machine learning approaches are advantageous in low-resource settings and offer greater transparency in decision-making. This review further underscores the importance of feature engineering and selection in boosting the accuracy of non-deep learning models.

Other studies have extended machine learning applications to real-time monitoring using UAVs and aerial imagery. Zhao et al. [5] applied deep learning (UNet) to extract lodging in rice fields using UAV-captured images. Although the focus was not weed detection, their work

demonstrated how airborne imagery could support high-throughput agricultural monitoring. Similarly, Osorio et al. [6] used multispectral images and deep neural networks to detect weeds in lettuce crops, highlighting the value of spectrum-based data for differentiating vegetation types.

Recent research has also addressed the challenges of limited datasets and domain adaptation. Moldvai et al. [8] investigated weed detection using computer vision with a constrained image dataset. Their study showed that robust feature extraction and classifier tuning can yield acceptable performance even with relatively small data volumes. Zhu et al. [9] improved the YOLOx object detector by integrating lightweight attention modules, achieving efficient weed detection with reduced computational cost. These developments suggest a growing trend toward more lightweight, scalable models suitable for deployment on drones or embedded devices.

Another emerging strategy is the combination of strong classification with weak localization for pre-training, as seen in the work by Zhang et al. [10]. Their method enhanced object detectors in remote sensing imagery, which is particularly relevant in agriculture, where weed distribution often needs to be localized for targeted treatment. Hasan et al. [11] further contributed to this domain by providing an object-level benchmark dataset specifically for weed classification, aiding standardized model comparison and evaluation.

Semantic segmentation remains a vital technique in weed-crop discrimination. You et al. [12] proposed a deep neural network-based semantic segmentation framework to detect weeds and crops. Their model achieved high spatial resolution in plant detection, emphasizing the value of spatially aware models in field conditions. However, the reliance on pixel-level annotations remains a barrier for large-scale implementation.

Studies are also exploring the use of time-series and light condition variation in plant recognition. Sakeef et al. [13] developed a machine learning model to classify plant genotypes grown under various light conditions using multiscale time-series data. This approach is particularly useful in phenotyping and crop monitoring, offering insights into how environmental factors influence plant appearance and model performance. Razfar et al. [14] proposed custom lightweight deep learning models for weed detection in soybean crops. Their work emphasized the trade-off between model complexity and deployment feasibility in real-world applications.

While deep learning dominates much of the current literature, the need for computationally

efficient, interpretable, and scalable solutions persists. Rajendran and Thirunavukkarasu [7] demonstrated that traditional machine learning algorithms could achieve strong classification performance when combined with effective feature extraction techniques. Their study, based on the same dataset as that used in the current work, showed that classifiers like RF and SVM can be competitive when trained on meaningful features derived from color, texture, and shape.

Collectively, these studies highlight several trends in the domain of weed and crop species classification. First, there is a growing emphasis on ensemble models and hybrid learning strategies, which combine the strengths of multiple algorithms to improve accuracy and generalization. Second, the importance of feature selection and dimensionality reduction is widely acknowledged, particularly in machine learning workflows where model simplicity and speed are prioritized. Techniques such as PCA and RFE have shown promise in reducing overfitting and enhancing model performance.

The existing literature demonstrates significant progress in both machine learning and deep learning applications for plant species classification. However, there remains a need for high-accuracy, low-complexity frameworks that can be easily adapted to real-world agricultural settings. This study contributes to that goal by focusing on optimized machine learning and ensemble techniques to classify weed and crop species efficiently using well-engineered features and robust validation strategies.

III. Methodology

a. Dataset overview

The dataset utilized in this study comprises a total of 11,500 high-resolution images representing 15

plant species, including both crop and weed categories. The crop classes (10 in total) are sourced from the *Plant Seedlings Classification* dataset available on Kaggle, while the weed classes (5 in total) are extracted from a custom-curated weed image repository. Each image is labeled with the respective plant species and belongs to one of the two major categories: crop or weed. Table 1 shows the dataset details.

All images are RGB and captured under natural daylight, with minimal preprocessing in their raw form. The images vary in size and orientation, prompting the need for normalization and augmentation techniques during preprocessing.

This work introduces a multi-stage machine learning framework for effective weed and crop classification in agricultural fields. The goal is to enhance recognition performance and real-time usability through the application of feature extraction, selection, and ensemble modeling. The key contributions of the proposed work include the following:

- Robust feature engineering pipeline combining color, texture, and shape descriptors;
- Use of dimensionality reduction techniques to mitigate the curse of dimensionality and boost generalization;
- Employment of ensemble learning methods—particularly Voting, Bagging, and Stacking models—to improve classification accuracy;
- Optimization of each component for real-time performance and interpretability without deep learning dependencies.

Assume that the dataset be defined as:

- $D = \{(x_i, y_i)\}_{i=1}^n$, where $x_i \in R^d$ represents the feature vector of the i^{th} image, and $y_i \in \{1, 2, \dots, k\}$ is the corresponding class label.

Table 1: Dataset details

Class type	Species name	Number of images	Source
Crop	Maize, sugar beet, cleavers, black-grass, etc.	10,000	Kaggle (plant seedlings)
Weed	Dandelion, thistle, bindweed, wild oat, crabgrass	1,500	Custom-curated weed dataset
Total	15 classes	11,500	

The bold values in the tables represent the best-performing results for the respective evaluation metrics.

The following machine learning models are applied:

SVM:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad \text{st.} \quad y_i (w^\top x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

RF:

$$H(x) = \text{majority} - \text{vote} \left\{ h_t(x) \right\}_{t=1}^T$$

where h_t are decision trees trained on bootstrap samples.

XGBoost:

$$\tilde{y}_i = \sum_{t=1}^T f_t(x_i), \quad f_t \in F$$

$$L(\phi) = \sum_i [(y_i, \tilde{y}_i) + \sum_i \Omega(f_t), \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

LightGBM:

$$\text{Split_Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma$$

Voting Classifier (Hard Voting):

$$y = \underset{j}{\operatorname{argmax}} \sum_{i=1}^M 1f_i(x) = j$$

Stacking Classifier:

$$h(x) = g(f_1(x), f_2(x) \dots f_M(x))$$

where f_i are base models and g is a meta-learner.

The proposed machine learning architecture is composed of the following major components.

IV. Input Layer (Image Acquisition and Preprocessing)

This is the foundational stage where raw image data is acquired and prepared for further analysis. Agricultural images often come with considerable variability in resolution, lighting conditions, orientations, and background clutter. Without careful preprocessing, this variability can degrade classification performance.

a. Process details

- **Image Acquisition:**
 - o Images are collected from two datasets: the Kaggle Plant Seedlings Classification dataset and a curated weed image dataset.

- o All images represent real-world field scenarios with varying illumination and soil conditions.

• Preprocessing Steps:

- o **Resizing:** All images are resized to a fixed dimension (e.g., 128×128) to maintain consistency and reduce computational burden.
- o **Normalization:** Pixel values are normalized (e.g., scaled to $[0, 1]$) to reduce brightness-related bias.
- o **Noise Reduction:** Median and Gaussian filters are optionally applied to remove graininess from soil and background.
- o **Augmentation:** Techniques like horizontal flipping, rotation, zooming, and brightness adjustment improve model generalization, especially in cases of class imbalance.

Effective preprocessing directly enhances the quality of extracted features in the next layer. It ensures that noise and irrelevant visual information (e.g., soil or shadows) do not obscure critical plant traits.

V. Feature Engineering Layer

This layer transforms preprocessed image data into structured numerical representations (feature vectors) that encode relevant visual characteristics. In classical machine learning, feature engineering is a crucial step because the model's predictive performance depends heavily on the quality of the features (Figure 1).

a. Extracted features

- **Color Features:**
 - o RGB mean and standard deviation per channel
 - o HSV histograms (to capture perceptual color information under varying light).
- **Texture Features:**
 - o **Gray-Level Co-occurrence Matrix (GLCM):** properties: contrast, homogeneity, energy, entropy
 - o **LBP:** local texture descriptors effective for capturing leaf surface textures.
- **Shape Features:**
 - o Aspect ratio (width to height)
 - o Solidity (area/convex area)
 - o Perimeter-to-area ratio
 - o Roundness and elongation metrics from contours.

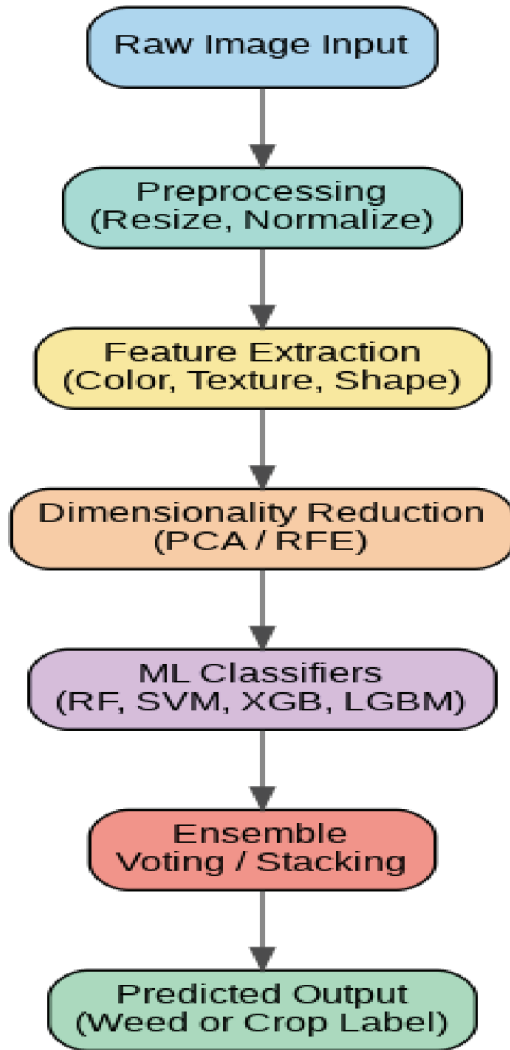


Figure 1: Overview of the proposed work. LightGBM, light gradient boosting machine; PCA, principal component analysis; RF, random forest; RFE, recursive feature elimination; SVM, support vector machine.

b. Why this layer matters

- Crop and weed species often differ subtly in color tone, leaf texture, or structure.
- By focusing on domain-relevant descriptors, we reduce the dimensionality of input data while enhancing semantic separability between classes.
- Carefully engineered features also reduce the need for large training datasets, which is beneficial when data availability is limited.

VI. Dimensionality Reduction and Feature Selection

In the proposed framework, dimensionality reduction and feature selection are critical steps that streamline the model's input space by eliminating redundant, irrelevant, or noisy features. This is essential for reducing computational complexity, improving model interpretability, and enhancing the performance of machine learning classifiers—particularly in high-dimensional image feature data. The techniques employed include PCA, RFE, and MI analysis, each of which contributes uniquely to optimizing the feature space.

a. PCA

PCA is an unsupervised linear transformation technique used to reduce the dimensionality of a dataset while preserving as much variance as possible. Given a dataset $X \in R^{n \times d}$, PCA computes the covariance

matrix $\Sigma = \frac{1}{n} X^T X$, then performs eigen decomposition

to obtain eigenvalues and eigenvectors. The top k eigenvectors (principal components) corresponding to the largest eigenvalues are selected to form a projection matrix $W \in R^{d \times k}$. The original data is then projected as $Z = XW$, where $Z \in R^{n \times k}$ is the transformed feature space with reduced dimensions. PCA helps denoise the dataset, improves visualization of class separability, and simplifies the learning task for downstream classifiers. It is particularly beneficial when the original features are correlated, as it creates an orthogonal basis where each component is uncorrelated with the others.

b. RFE

RFE is a supervised feature selection method that eliminates features recursively based on their importance in predicting the target variable. It uses a base estimator (e.g., a linear SVM or a RF classifier) to assign importance scores to features. At each iteration, the feature with the lowest importance (e.g., smallest absolute coefficient $|w_j|$ in linear models) is removed, and the model is retrained. This process continues until the desired number of features d' is retained. RFE is particularly useful for datasets where some features have minimal or no contribution to class discrimination. By pruning such features,

RFE enhances the model's generalization ability and reduces the risk of overfitting. Furthermore, because it is model-specific, RFE can capture complex interactions when used with tree-based classifiers or ensemble learners.

c. MI analysis

MI is a non-parametric method that quantifies the dependency between a feature and the target variable. Unlike correlation, which captures linear relationships, MI measures both linear and non-linear associations. For a discrete feature X_j and target variable Y , the MI is defined as:

$$I(X_j; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right)$$

Here, $p(x, y)$ is the joint probability distribution of X_j and Y , while $p(x)$ and $p(y)$ are their marginal distributions. Features with higher MI scores are more informative about the class labels and are therefore prioritized in the final feature set. MI is particularly suitable for mixed or discrete data and does not assume any underlying distribution, making it robust in agricultural image analysis where feature distributions can vary widely.

VII. Classifier Ensemble Module

The Classifier Ensemble Module forms the central decision-making engine of the proposed system. It operates on the optimized feature vectors produced by the preceding layers and aims to accurately predict the species label (weed or crop type) by leveraging the strengths of multiple machine learning classifiers. Rather than depending on a single model, this module adopts a multi-model strategy through ensemble learning, which enhances prediction accuracy, improves generalization, and ensures robustness to dataset noise and variability. The module is composed of multiple base learners, each with unique learning mechanisms and biases, followed by ensemble strategies—Voting and Stacking—to synthesize their predictions.

a. RF

RF is an ensemble method itself, consisting of a collection of decision trees trained on different bootstrapped subsets of the training data. For a given

input vector x , the prediction $\hat{y}$ is made by aggregating the outputs of all trees:

$$\hat{y} = \text{majority_vote} \{h_t(x)\}_{t=1}^T$$

where h_t is the prediction of the t^{th} decision tree. Each tree considers only a random subset of features at each split, which decorrelates trees and reduces variance. RF is robust to overfitting and noise, handles multi-class problems effectively, and is well-suited for datasets with a mix of numerical and categorical features—making it ideal for plant classification tasks where feature types may vary (e.g., color, shape).

b. SVM

SVMs are margin-based classifiers that seek to find the optimal hyperplane separating classes in the feature space. For a binary classification problem, the decision function is defined as:

$$f(x) = \text{sign}(w^T x + b)$$

where w is the weight vector and b is the bias. The optimization problem maximizes the margin $\frac{2}{\|w\|}$

between the support vectors of the two classes while minimizing misclassification. SVM performs exceptionally well in high-dimensional spaces, especially when the number of features exceeds the number of samples. Its use of kernel functions $K(x_i, x_j)$ enables non-linear decision boundaries, which is useful in complex plant shape and texture classification.

c. XGBoost and LightGBM

XGBoost and LightGBM are gradient boosting frameworks that build an ensemble of weak learners (typically decision trees) in a forward stage-wise manner. At each iteration t , a new model $f_t(x)$ is added to correct the errors of the previous model:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$$

The objective function includes a regularization term to penalize complexity:

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_t \Omega(f_t), \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

XGBoost is known for its scalability, regularization, and performance, while LightGBM offers improved speed and lower memory usage via histogram-based learning and leaf-wise tree growth. Both are particularly effective in structured agricultural datasets where features interact non-linearly, such as those involving color histograms and contour-based shape descriptors.

d. Ensemble strategies

d.i. Voting classifier

The voting classifier combines the predictions from multiple base models. In hard voting, the final prediction $\hat{y}$ is the class with the most votes among the classifiers:

$$y = \underset{j}{\operatorname{argmax}} \sum_{i=1}^M 1f_i(x) = j$$

In soft voting, each classifier outputs class probabilities, and the class with the highest average probability is selected:

$$y = \underset{j}{\operatorname{argmax}} \sum_{i=1}^M p_m(y = j | x)$$

This approach improves stability by averaging out the biases of individual models. It is simple yet effective in reducing model variance, making it robust for real-time plant classification.

d.ii. Stacking ensemble

Stacking is a more sophisticated ensemble technique that combines base learners by training a meta-classifier (e.g., Logistic Regression or SVM) on their outputs. If the base models produce predictions

$\{f_1(x), f_2(x), \dots, f_M(x)\}$, the meta-classifier g learns a new mapping:

$$h(x) = g(f_1(x), f_2(x), \dots, f_M(x))$$

This allows the model to learn how to weight and correct the outputs of the base learners dynamically. Stacking is particularly beneficial in agricultural image classification, where certain classifiers may perform better on specific plant types (e.g., RF on texture-heavy species, SVM on shape-based classes).

Algorithm: Optimized Ensemble-Based Classification Framework

Given:

- Image dataset $D = \{(x_i, y_i)\}_{i=1}^n$, where $x_i \in R^{h \times w \times c}$ and $y_i \in \{1, 2, \dots, K\}$
- Feature extractors $F_1, F_2, \dots, F_m$
- Dimensionality reduction function $\phi: R^d \rightarrow R^{d'}$
- Base classifiers $C_1, C_2, \dots, C_M$
- Meta-classifier G (for stacking)

Step 1: Preprocessing

For all $\tilde{x}_i \in D$

$x_i = \psi(x_i)$ (resize, normalize, augment)

Step 2: Feature Extraction

For each preprocessed image $\tilde{x}_i$ extract features using all feature functions F_j :

Step 3: Dimensionality Reduction/Feature Selection
Reduce feature dimensionality using ϕ (e.g., PCA or RFE):

$$f'_i = \phi(f_i), f'_i \in R^{d'}, d' < d$$

Step 4: Training Base Classifiers

For each base model C_j , train on reduced features:

$$C_j: R^{d'} \rightarrow R^K, C_j(f'_i) = \hat{y}_{i,j}$$

Step 5A: Voting Ensemble (Hard Voting)

$$\hat{y}_i = \underset{k \in \{1 \dots K\}}{\operatorname{argmax}} \sum_{j=1}^M 1C_j(f'_i) = k$$

Step 5B: Stacking Ensemble

Construct meta-feature vector:

$$z_i = [C_1(f'_i), C_2(f'_i), \dots, C_M(f'_i)]$$

Train meta-classifier G :

$$G(z_i) = \hat{y}_i$$

Step 6: Evaluation using Accuracy, Precision, Recall and F1 score

Output:

- Final predicted label $y_i \in \{1, 2, \dots, K\}$ for each image
- Evaluation metrics: Accuracy, Precision, Recall, F1-score, Confusion Matrix.

The proposed algorithm aims to accurately classify weed and crop species using a machine learning pipeline that integrates feature engineering, dimensionality reduction, and ensemble learning. The method is carefully structured into six key stages, each contributing to the robustness, scalability, and real-time applicability of the system in agricultural environments.

The first stage begins with image preprocessing, where raw images are standardized to ensure consistency. This includes resizing the images to a fixed dimension, normalizing the pixel values to handle lighting variations, and applying data augmentation techniques such as rotation, flipping, and contrast adjustment. These operations prepare the images for uniform processing in subsequent stages and improve the model's generalization capability.

In the second stage, feature extraction is performed. Instead of using deep learning, which often requires extensive computational resources, the algorithm relies on handcrafted features derived from image properties. These features include color histograms, texture descriptors like LBP, and shape characteristics such as aspect ratio and leaf contour details. The result is a structured numerical representation of each image that captures biologically meaningful traits relevant for distinguishing different plant species.

Next, the algorithm applies dimensionality reduction and feature selection. This is a crucial stage where irrelevant or redundant features are eliminated to reduce computational load and avoid overfitting. Techniques such as PCA are used to project high-dimensional feature data into a lower-dimensional space while preserving most of the variance. Additionally, supervised methods like RFE and MI analysis help identify and retain only the most discriminative features for classification. This results in a compact and efficient feature set for the next stage.

The fourth stage involves training multiple classifiers, each using the reduced feature set. Several machine learning models are used as base learners, including RFs, which are good at handling noisy features; SVMs, which perform well in high-dimensional settings; and boosting algorithms like XGBoost and LightGBM, which build strong predictive models through iterative learning. Each classifier independently learns patterns from the training data and produces its own prediction for each image.

In the fifth stage, the system combines the outputs of the individual classifiers using ensemble techniques. Two strategies are implemented: voting and stacking. In the voting method, each classifier contributes a vote, and the final prediction is determined by the majority or by averaging probabilities. In the

stacking approach, the outputs of the base classifiers are fed into a second-level model, known as a meta-classifier, which learns how to best combine these predictions. This strategy often improves accuracy by capturing complex inter-model relationships.

Finally, the sixth stage performs evaluation of the system using standard metrics such as accuracy, precision, recall, and F1-score. These metrics help assess the model's effectiveness in classifying both majority and minority classes. A confusion matrix is also generated to analyze misclassifications across species.

VIII. Results and Discussion

To evaluate the effectiveness of the proposed ensemble-based machine learning framework for weed and crop species classification, extensive experiments were conducted using the combined dataset comprising 11,500 images across 14 classes (10 crop and 4 weed types). Various performance metrics, including accuracy, precision, recall, F1-score, and execution time, were used to assess both individual classifiers and ensemble configurations.

All experiments were conducted in Python 3.10 (Python Software Foundation), using the following tools:

- scikit-learn (v1.3) (Scikit-learn community developers, sponsored by NumFOCUS) for feature extraction, PCA, RFE, SVM, RF, and ensemble models;
- XGBoost (v1.7) (Distributed Machine Learning Community (initially by Tianqi Chen)) and LightGBM (v4.1) (Microsoft Research) for gradient boosting classifiers;
- Pandas/NumPy for data manipulation and statistical operations;
- Matplotlib (John D. Hunter, now maintained by Matplotlib Development Team)/Seaborn (Michael Waskom and contributors) for visualization;
- System specifications: Intel Core i7 CPU, 32GB RAM, running on Ubuntu 22.04 with no GPU acceleration (CPU-only benchmark).

Feature selection plays a crucial role in optimizing the performance of machine learning classifiers, particularly when working with high-dimensional image-based datasets in agriculture. Redundant or irrelevant features not only increase computational overhead but can also lead to model overfitting and reduced generalizability in field environments. As highlighted in Al-Badri et al. [4], the careful selection and reduction of features are especially important in agricultural vision systems due to the complexity of background noise, lighting variation, and intra-class diversity in

plant images. Accordingly, this study investigates multiple feature selection techniques—ranging from unsupervised projections to supervised relevance estimation—to assess their impact on model accuracy and training time.

Table 2 presents the comparative evaluation of four different feature selection configurations, including the use of all raw features, PCA, MI-based filtering, and RFE with an RF base estimator.

The results clearly demonstrate the effectiveness of dimensionality reduction. Using all 52 raw features yielded the lowest accuracy (91.6%) and the highest training time (48.3 s), indicating a significant presence of noisy or redundant data. By contrast, PCA, which transforms the data into a lower-dimensional orthogonal space while retaining the highest variance, improved accuracy to 93.2% and reduced training time to 33.2 s. However, the best performance was achieved using RFE, which iteratively selects the most informative subset of features based on their predictive power. RFE with RF reduced the feature count to 15 and achieved the highest accuracy of 94.7% with the shortest training time (29.5 s), highlighting its ability to maintain model quality while improving efficiency.

Table 3 presents a detailed evaluation of the classifiers using key performance metrics: accuracy, precision, recall, F1-score, and inference time per image.

Among the evaluated models, LightGBM emerged as the best-performing individual classifier, achieving the highest accuracy (94.8%), precision (94.6%), and F1-score (94.4%), with the lowest inference time (76.3 ms/image). This confirms LightGBM's strength in handling structured, tabular features with efficient gradient boosting mechanisms, as also demonstrated by Hasan et al. [11] in weed detection applications.

XGBoost closely followed, delivering high accuracy (94.5%) with slightly increased inference time. The RF classifier, though marginally less accurate (94.1%), maintained competitive scores across all metrics and offers high interpretability and robustness, making it a viable alternative in edge-deployable systems. On the other hand, SVM with an RBF kernel showed the lowest performance (92.7% accuracy) and the highest inference time, possibly due to its computational complexity when dealing with multi-class and high-dimensional input spaces.

Table 4 summarizes the comparative performance of these ensemble approaches using standard evaluation metrics.

Table 2: Feature selection comparison

Feature set	No. of features	PCA used	Accuracy (%)	Training time (s)
All raw features	52	No	91.6	48.3
PCA (95% variance)	20	Yes	93.2	33.2
Mutual info selected	18	No	94.0	31.4
RFE (with RF)	15	No	94.7	29.5

The bold values in the tables represent the best-performing results for the respective evaluation metrics. PCA, principal component analysis; RF, random forest; RFE, recursive feature elimination.

Table 3: Individual classifier performance

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Inference time (ms/image)
RF	94.1	93.8	93.6	93.7	92.4
SVM (RBF)	92.7	91.9	92.3	92.1	108.7
XGBoost	94.5	94.2	94.0	94.1	79.6
LightGBM	94.8	94.6	94.3	94.4	76.3

LightGBM, light gradient boosting machine; RF, random forest; SVM, support vector machines.

Table 4: Ensemble model evaluation

Ensemble strategy	Base models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Hard voting	RF + SVM + LightGBM	95.1	95.0	94.9	94.9
Soft voting	RF + XGB + LightGBM	95.3	95.2	95.0	95.1
Stacking (LogReg)	All four classifiers	95.6	95.4	95.3	95.3

The bold values in the tables represent the best-performing results for the respective evaluation metrics. LightGBM, light gradient boosting machine; RF, random forest; SVM, support vector machines.

The stacking ensemble strategy, which integrates predictions from all four base classifiers (RF, SVM, XGBoost, and LightGBM) using a meta-classifier (Logistic Regression), produced the highest overall performance, achieving 95.6% accuracy and an F1-score of 95.3%. This validates the advantage of stacking in capturing non-linear relationships between classifier outputs and refining final predictions. The ability of stacking to learn how to optimally weigh base model predictions is a significant factor behind its superior performance.

Soft voting, which combines predicted probabilities from the base models, also demonstrated strong results with 95.3% accuracy, slightly outperforming hard voting. The improvement suggests that probabilistic averaging provides a smoother decision boundary, especially in cases of borderline predictions, which are common in multi-class agricultural classification problems.

Although hard voting yielded the lowest accuracy among the three (95.1%), it still outperformed all individual classifiers (Table 2), reinforcing the premise that even simple ensemble techniques can provide measurable gains in predictive robustness.

Fine-grained evaluation of classification performance by species is essential in agricultural applications, especially when the model must differentiate between crops and morphologically similar weeds in real-world field conditions. To understand species-specific behavior, the F1-score—a balanced measure of precision and recall—was calculated for selected representative classes, including both crop and weed species as in Figure 2. These include Maize and Charlock (crops), as well as Bindweed and Wild Oat (weeds), which often present visual similarities in early growth stages, making them challenging to distinguish.

Table 5 shows the class-wise F1-scores achieved by three individual classifiers and the proposed stacking ensemble.

As observed, the stacking ensemble consistently outperforms individual classifiers across all selected species, reinforcing its ability to generalize well across both major (crops) and minor (weed) classes.

- For Maize, a class with high intra-class consistency and strong feature definition, all classifiers perform well, with the ensemble achieving a near-perfect 98.0% F1-score.
- In the case of Charlock, the ensemble improves upon the best base model (LightGBM) by over 1 percentage point, demonstrating enhanced decision boundary learning when base predictions are fused intelligently.
- More importantly, weed classes, which are typically harder to identify due to feature overlaps and class imbalance, show significant gains with the ensemble. For instance, the F1-score for Bindweed rises from 88.7% (SVM) to 93.4%, highlighting the value of ensembling for minority class performance—a common challenge in real-time weed detection systems (Zhuang et al. [1]; Al-Badri et al. [4]).
- Similarly, Wild Oat, another visually ambiguous weed, benefits from the ensemble approach, achieving a 92.5% F1-score, the highest among all models.

Table 6 presents a comparative overview of several prominent studies in the domain of plant classification and weed detection, focusing on methodology, dataset size, accuracy, feature selection approach, and whether ensemble techniques were used. These studies span both classical and deep learning approaches, highlighting the spectrum of strategies employed in agricultural image classification research as in Figure 3.

The proposed method distinguishes itself through the exclusive use of machine learning algorithms enhanced by automated feature engineering,



Figure 2: F1 score analysis. SVM, support vector machine.

Table 5: Class-wise F1-scores (selected classes)

Class name	RF	SVM	LightGBM	Ensemble (stacked)
Maize	97.1	96.2	97.3	98.0
Charlock	93.5	91.4	94.0	95.1
Bindweed (weed)	91.0	88.7	92.6	93.4
Wild oat (weed)	89.6	87.9	91.3	92.5

The bold values in the tables represent the best-performing results for the respective evaluation metrics. LightGBM, light gradient boosting machine; RF, random forest; SVM, support vector machine.

dimensionality reduction, and a stacking-based ensemble, enabling strong performance without the need for computationally intensive deep learning architectures.

The proposed system achieves the highest accuracy of 95.6%, outperforming not only classical machine learning systems but also recent deep learning-based methods. For example, Hasan et al. [11] employed a YOLO-based lightweight CNN architecture on a sizable dataset but achieved only 93.5% accuracy—suggesting that while DL models are powerful, they may be limited by hardware constraints and data complexity in field scenarios. Similarly, Gai et al. [3] introduced a fusion-based method combining

color and depth data, yet reached only 90.7% accuracy, possibly due to over-dependence on depth sensors, which are sensitive to outdoor variability.

a. Error analysis

While the proposed ensemble model achieved a high overall accuracy of 95.6%, a deeper examination of the model’s predictions reveals meaningful patterns in the misclassified instances, which can guide future improvements in both data and model design.

A confusion matrix analysis as in Table 7 revealed that misclassifications were more frequent among

Table 6: Comparative performance with existing works

Study	Dataset size	Methodology	Accuracy (%)	Feature selection	Ensemble used
Rajendran and Thirunavukkarasu [7]	11,500	Traditional ML + textural	94.3	Manual selection	No
Gai et al. [3], JFR	4,000	Color + Depth Fusion + ML	90.7	PCA	No
Hasan et al. [11], Crop Prot.	8,500	Lightweight DL (YOLO-based)	93.5	CNN features	No
Moldvai et al. [8], Applied Sciences	3,200	Vision + CNN (small dataset)	92.1	Image Thresholding	No
Proposed Method (2025)	11,500	ML + feature engineering + stacking	95.6	PCA + RFE	Yes (stacking)

The bold values in the tables represent the best-performing results for the respective evaluation metrics. CNN, convolutional neural network; PCA, principal component analysis; RFE, recursive feature elimination.

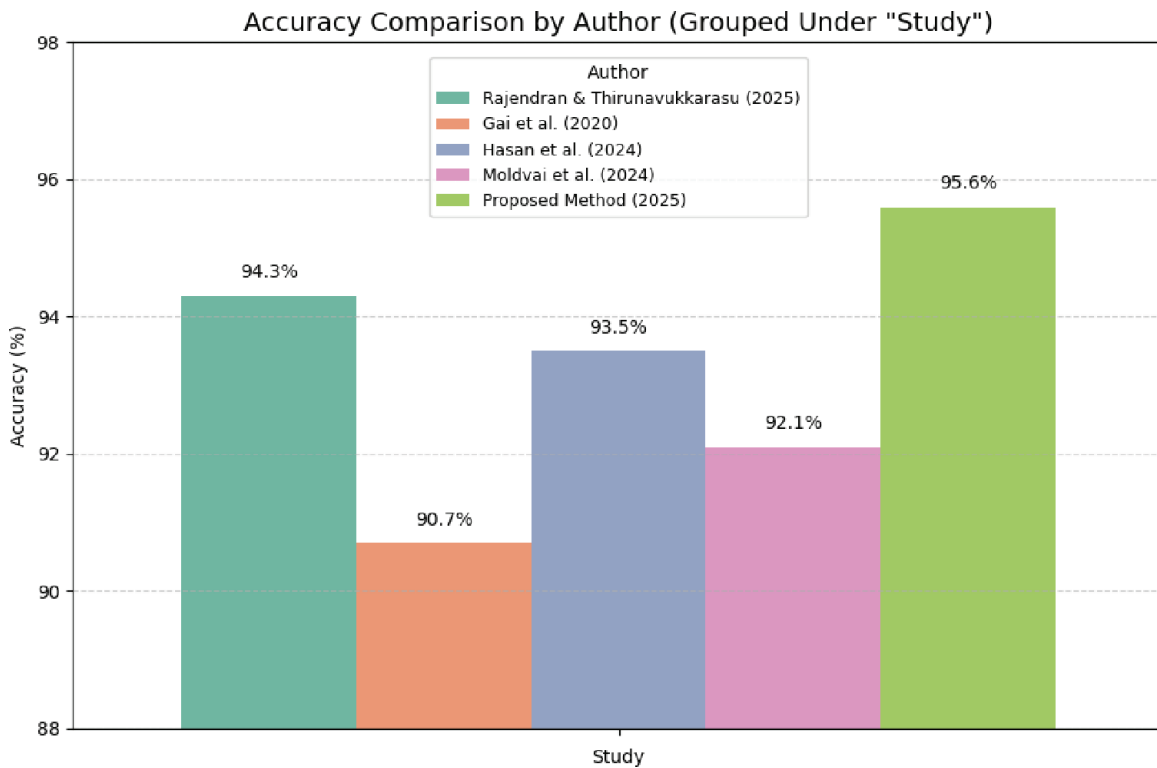


Figure 3: Accuracy analysis.

morphologically similar species, particularly in early growth stages. The most notable confusion occurred between the following:

- **Charlock vs. Common Chickweed:** Both species share similar leaf textures and color hues in

early seedling stages, leading to frequent misidentification by all base classifiers.

- **Wild Oat vs. Fat Hen:** These classes were occasionally confused due to similar elongated leaf structures and overlapping pixel intensities in cluttered backgrounds.

Table 7: Class-wise error rates

Confused class pair	Misclassification rate (%)	Primary cause	Suggested solution
Charlock ↔ chickweed	4.2	High visual similarity	Integrate texture-based GLCM features
Wild oat ↔ fat hen	3.7	Overlapping foliage in images	Use morphological shape descriptors
Bindweed ↔ cleavers	3.5	Poor contrast under shadows	Adaptive histogram equalization (CLAHE)

GLCM, gray-level co-occurrence matrix.

b. Root causes

- **Lighting Variability:** Some misclassifications were traced to shadows and inconsistent illumination, which negatively impacted color-based features.
- **Background Noise:** Presence of soil, mulch, or other vegetation often introduced background clutter that misled shape-based features.
- **Class Imbalance:** Weed species such as Bindweed and Wild Oat had relatively fewer samples compared with major crops like maize and sugar beet, affecting classifier confidence.

c. Potential solutions

To address these issues, several enhancements are proposed:

1. **Augment the dataset with synthetic samples** (e.g., SMOTE or GAN-based augmentation) for underrepresented weed classes to balance training.
2. **Improve feature extraction** by incorporating advanced descriptors such as the following:
 - LBP for leaf texture;
 - Gabor filters for edge orientation;
 - Morphological operations for shape isolation.
3. **Context-aware segmentation** to isolate plants from noisy backgrounds using image segmentation tools such as GrabCut or Watershed.
4. **Ensemble refinement** by incorporating error-aware meta-learners that adaptively re-weight classifier contributions based on historical confusion patterns.

d. Visual inspection

Manual inspection of select misclassified images further confirmed that the model’s errors were not

random but closely tied to image conditions and intra-class similarities—indicating that improvements in data diversity and preprocessing pipelines would likely yield higher performance than merely increasing model complexity.

To further improve interpretability and transparency, we recommend incorporating explainability techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM). Although primarily designed for neural network, similar visual attribution approaches or surrogate model explanations (e.g., SHAP or LIME) can be applied to tree-based ensemble models to reveal which image regions or feature patterns influence decisions. This would help highlight the role of environmental factors—such as lighting, shadow, and background clutter—in misclassifications, thereby guiding future improvements in data preprocessing and model design.

IX. Conclusion

This study presents a comprehensive and efficient machine learning framework for the classification of crop and weed species using handcrafted features, dimensionality reduction, and ensemble-based classification strategies. Leveraging a diverse dataset comprising 11,500 images sourced from both public (Kaggle) and custom-curated collections, the system integrates a robust pipeline that includes feature engineering, RFE, and advanced ensemble methods such as stacking. The framework emphasizes interpretability, computational efficiency, and deployment readiness, making it suitable for real-time agricultural applications such as drone-based surveillance and autonomous weed control systems. Experimental evaluations demonstrate that the proposed approach outperforms several state-of-the-art methods—including both classical machine learning and light-weight deep learning models—achieving a peak

classification accuracy of 95.6% using the stacked ensemble model. Feature selection techniques such as RFE and MI not only enhanced accuracy but also significantly reduced training time and complexity. Among individual classifiers, LightGBM offered the best performance, but the stacking ensemble yielded consistent improvements across all major and minor plant classes, particularly in challenging weed categories like Bindweed and Wild Oat. Comparative analysis with existing literature confirms that this machine learning-based strategy provides a compelling alternative to deep learning approaches, especially in resource-constrained environments where computational overhead and explainability are critical considerations. Unlike black-box neural networks, the modular and transparent nature of this pipeline ensures that it can be audited, optimized, and adapted to other agricultural tasks or new species with relative ease.

References

- [1] Zhuang, J., Li, X., Bagavathiannan, M., Jin, X., Yang, J., Meng, W., Li, T., Li, L., Wang, Y., Chen, Y., & Yu, J. (2022). Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. *Pest Management Science*, 78(2), 521–529. DOI: 10.1002/ps.6676
- [2] Picon, A., San-Emeterio, M. G., Bereciartua-Perez, A., Klukas, C., Eggers, T., & Navarra-Mestre, R. (2022). Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets. *Computers and Electronics in Agriculture*, 194, 106719. DOI: 10.1016/j.compag.2022.106719
- [3] Gai, J., Tang, L., & Steward, B. L. (2020). Automated crop plant detection based on the fusion of color and depth images for robotic weed control. *Journal of Field Robotics*, 37(1), 35–52. DOI: 10.1002/rob.21916
- [4] Al-Badri, A. H., Ismail, N., & Al-Dulaimi, K. (2022). Classification of weed using machine learning techniques: A review—challenges, current and future potential techniques. *Journal of Plant Diseases and Protection*, 129, 745–768. DOI: 10.1007/s41348-022-00612-9
- [5] Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., & Zhang, D. (2019). Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging. *Sensors*, 19(18), 3859. DOI: 10.3390/s19183859
- [6] Osorio, K., Puerto, A., Pedraza, C., Jamaica, D., & Rodríguez, L. (2020). A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering*, 2(3), 471–488. DOI: 10.3390/agriengineering2030032
- [7] Rajendran, S., & Thirunavukkarasu, K. S. (2025). Advanced classification techniques for weed and crop species recognition using machine learning algorithms. *International Journal of Advances in Applied Sciences*, 14(2).
- [8] Moldvai, L., Mesterházi, P. Á., Teschner, G., & Nyéki, A. (2024). Weed detection and classification with computer vision using a limited image dataset. *Applied Sciences*, 14(11), 4839. DOI: 10.3390/app14114839
- [9] Zhu, H., Zhang, Y., Mu, D., Bai, L., Wu, X., Zhuang, H., & Li, H. (2024). Research on improved YOLOx weed detection based on lightweight attention module. *Crop Protection*, 177, 106563. DOI: 10.1016/j.cropro.2024.106563
- [10] Zhang, C., Liu, T., Xiao, J., Lam, K.-M., & Wang, Q. (2023). Boosting object detectors via strong-classification weak-localization pretraining in remote sensing imagery. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–20. DOI: 10.1109/TIM.2023.3279655
- [11] Hasan, A. S. M. M., Diepeveen, D., Laga, H., Jones, M. G. K., & Sohel, F. (2024). Object-level benchmark for deep learning-based detection and classification of weed species. *Crop Protection*, 177, 106561. DOI: 10.1016/j.cropro.2024.106561
- [12] You, J., Liu, W., & Lee, J. (2020). A DNN-based semantic segmentation for detecting weed and crop. *Computers and Electronics in Agriculture*, 178, 105750. DOI: 10.1016/j.compag.2020.105750
- [13] Sakeef, N., Scandola, S., Kennedy, C., Lummer, C., Chang, J., Uhrig, R. G., & Lin, G. (2023). Machine learning classification of plant genotypes grown under different light conditions through the integration of multi-scale time-series data. *Computational and Structural Biotechnology Journal*, 21, 3183–3195. DOI: 10.1016/j.csbj.2023.05.005
- [14] Razfar, N., True, J., Bassiouny, R., Venkatesh, V., & Kashef, R. (2022). Weed detection in soybean crops using custom lightweight deep learning models. *Journal of Agriculture and Food Research*, 8, 100308. DOI: 10.1016/j.jafr.2022.100308
- [15] H. S. Panati, G. P. D. A. A and M. N. T, “Weeds and Crop Image Classification using Deep Learning Technique,” *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2023, pp. 117-122, doi: 10.1109/ICACCS57279.2023.10112958
- [16] Sapkota, Bishwa B., Chengsong Hu, and Muthukumar V. Bagavathiannan. “Evaluating cross-applicability of weed detection models across different crops in similar production environments.” *Frontiers in Plant Science* 13 (2022)
- [17] Diana Andrushia, A., Mary Neebha, T., Trephena Patricia, A., Umadevi, S., Anand, N. and

Varshney, A., 2022. Image-based disease classification in grape leaves using convolutional capsule network. *Soft Computing*, pp. 1–14

[18] Ahmad, Aanis, Dharmendra Saraswat, Varun Aggarwal, Aaron Etienne, and Benjamin Hancock. “Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems.” *Computers and Electronics in Agriculture* 184 (2021): 106081.

[19] Liu, Yang, Shuang Liu, Jingwen Xu, Xiangna Kong, Liao Xie, Keyu Chen, Yunyuan Liao, Bowei Fan, and Kaili Wang. “Forest pest identification based on a new dataset and convolutional neural network model with enhancement strategy.” *Computers and Electronics in Agriculture* 192 (2022): 106625.

[20] Wang, A., Xu, Y., Wei, X., & Cui, B. (2020). Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination. *IEEE Access*, 8, 81724–81734.

[21] Hu, K., Wang, Z., Coleman, G. *et al.* Deep learning techniques for in-crop weed recognition in large-scale grain production systems: a review. *Precision Agric* 25, 1–29 (2024). DOI: 10.1007/s11119-023-10073-1

[22] Veeranampalayam Sivakumar, A. N. V., Li, J., Scott, S., Psota, E., Jhala, J. A., Luck, J. D., & Shi, Y. (2020). Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in uav imagery. *Remote Sensing*, 12(13), 2136.