Weed Identification with Machine learning: A Path to Smart Farming

Zoha Attar[1*] and Dr.Meenakshi Thalor[2]

[1*]Information Technology, AISSMS Institute of Information Technology, , Pune, Maharashtra, India.
[2]Information Technology, AISSMS Institute of Information Technology,Pune,Maharashtra, India.

Abstract

Agriculture plays a significant role in contributing to the Indian economy. Several obstacles stand in the way of its potential, most notably the ubiquitous problem of weed infestation. Weeds affect crop quality and yield by competing with crops for vital nutrients. This proposed approach to weed detection deviates from traditional techniques by integrating the use of Detection Transformer (DETR) for object recognition and Vision Transformer for feature extraction. This combination of techniques, in contrast to CNN-based models, seeks to address shortcomings in spatial resolution, contextual knowledge, and scale invariance. It aims to boost weed identification accuracy, thus contributing towards the agriculture sector's sustainability.

Keywords: Vision transformers, Detection Transformers, Precision Agriculture

1 Introduction

Agriculture forms the foundation of human civilization, feeding populations and powering economic growth for hundreds of years. The agriculture sector contributes about 17 percent GDP to the Indian economy. Traditional farming practices face problems of declining profits and productivity. Hence, there is a need for sustainable agriculture practices. Precision agriculture, a term coined in recent decades, addresses the challenges traditional agriculture faces, such as unpredictable weather, outbreaks of weeds, pests, and diseases. It uses technology and data along with agriculture science to monitor and manage crops, enabling farmers to make data-driven real-time decisions on the use of resources to optimize crop production for better yields. Weeds (unwanted plants) are often classified as invasive species and compete with crops for essential nutrients. They also have a major impact on crop yield quality and reduced productivity. Traditional weed management practices, such as manual weed detection, can be time consuming, labor intensive, and costly and can lead to incomplete removal of weeds. Using technology to detect weeds can help farmers identify and target areas infested with weeds more accurately and quickly. In agriculture, the usual method for plant detection involves using segmentation to separate vegetation from the background and then distinguishing weeds from crops. Although multi spectral data can be used to differentiate between vegetation and the background, crops and weeds are often challenging to differentiate using only spectral information because they have similar features. Many studies on weed classification or detection have used models such as AlexNet, VGG, Google Net, ResNet, and Inception-V3. However, the image classification task does not provide information about the location of the weed and is incapable of identifying numerous instances within one image. Object detection, on the other hand, is capable of providing the location of different weeds inside an image using bounding boxes. So, the advanced transformer architecture is proposed, which was originally introduced in natural language processing (NLP) tasks. However, it has recently found applications in computer vision as well. The Transformer self-attention mechanism (Vaswani et al.) can capture relationships between different image regions, allowing it to effectively capture spatial dependencies in images. This has been shown to surpass state-of-the-art models in object detection and image classification tasks. In particular, the Vision Transformer (ViT) has shown great promise in image classification, achieving high accuracy on various benchmark datasets. ViT divides the input image into a sequence of patches. These patches are processed through a series of Transformer blocks before passing the output through a linear classifier (Dosovitskiy et al.). DETR is a transformer-based encoder-decoder framework for object detection. It combines object queries, which represent the types and locations of objects being detected, with a set of learned feature maps to directly output the final set of object detections.

2 Related Work

Convolution Neural Networks (CNNs) are popular in image-based applications for their ability to learn spatial and hierarchical features. The latest developments in computer vision technology and machine learning algorithms have revolutionized weed detection in crop fields. Traditional methods, reliant on manual feature extraction and machine learning classifiers, are being supplanted by deep learning approaches, which can automatically extract complex spatial and spectral features from images, leading to more accurate and efficient weed detection. However, despite the progress in weed detection techniques, challenges persist, including occlusion, variability in lighting conditions, and the inherent similarities between crop plants and weeds. Addressing these challenges requires developing robust and scalable algorithms for weed identification capable of operating under diverse field condi-tions. By examining the latest advancements in both traditional machine learning methods and deep learning algorithms, this study intends to provide insights into the current state-of-the-art, identify challenges, and propose future research directions in the field of weed detection. Through empirical evaluation and comparison of different approaches, this research aims to contribute to developing more effective, sustainable practices in agriculture.

Punithavathi et al. came up with an approach CVDL-WDC. The technique is based on primarily two major elements, namely multi scale Faster RCNN based object detection and optimal extreme learning machine (ELM) based weed classification. The model was trained and tested over benchmark dataset of crops and weeds. It has achieved highest accuracy of 98.33 percent at training/testing ratio of 80:20. When compared with other approaches like GW-GFD (GaborWavelet), GLCM (GrayLevel Co-occurence Matrix), and FCN- RCWD (Fully Convolutional Network) which gave accuracies of 93.75, 91.60, and 93.88 percent correspondingly, CVDL-WDC has performed significantly better. Pauline Ong, Kiat Soon Teo and Chee Kiong Sia compared conventional ML model Random Forest with CNN (AlexNet) based approach. This study aimed to detect weeds in Chinese cabbage plantation. A total of 700 images were collected manually using drones with attached cameras. These acquired images were pre-processed and segmented using Simple Linear Iterative Clustering (SLIC). SLIC hyper-parameters: the number of super pixels, K and compactness,. Unlike AlexNet, RF requires a feature extraction step before classification, hence an addition step using Local Binary Pattern (LBP) for extraction was used. Results show that CNN has outperformed RF with a 92.41 percent average accuracy. CNN misclassified weed classes due its resemblance with cabbage leaves at early stage of growth.

Singh et al. emphasized on lack of experimentation in CNN architectures. They used the standard Mendeley Data of 15,336 images. 15 distinct CNN architectures were proposed based on CL-The number of CNN Layers in an architecture, AL-The number of ANN Layers in an architecture, L1-Level 1 regularizations, L2-Level 2 regularizations, BN-Batch Normalization and a few more factors. To select a potentially useful model, Objective Function Value (OFV) which was calculated by using 3 parameters that were Maximum Validation Accuracy (MVA), Least Validation Cross-Entropy Loss (LVCEL) and Training Time (TT). The study compared the selected model with transfer learning approach on Xception, VGG16, and VGG19, and these performed better than the custom CNN model. Since CNN requires is significantly lesser time for training, there is scope for future advancements in the CNN model by usage of better datasets and more parameters.

Sanjay et al. tried segmentation along with vision transformers. The segments were ex-tracted from 400 aerial images. These images were captured by an unmanned aerial vehicle (UAV). The SLIC algorithm, which was implemented in the Pynovisao software, was used to pre-process them. The loss function used for the ViT is the Sparse Categorical Cross En-tropy (SCCE). The proposed model has attained an accuracy of 99.18 percent on the train-ing dataset of images, and 99.2 percent accuracy on the validation set. Data was classified in as "Broadleaf", "Grass", "Soil", "Soybean" and which had 100, 93,98,99 percentage of accuracies respectively. The study of L Uday Kumar Reddy, S Rohitharun, and S Sujana (CNN-based AlexNet Weed Detection on

DeepWeeds Dataset), pointed out a major research gap in weed identification. These gaps emphasized the lack of classification and effective methods for extracting features. To solve this issue, AlexNet architecture has been utilized. DeepWeeds and weed crop data sets have been used in the study. The proposed methodology was suc-cessful in identifying weeds in a particular patch of field with an average accuracy of 96 percent.

N.P, S.K., S. M. Ganesh, G.B. Prakash Yadav, T.S. Suneeth and K.M. Kumar used the CottonWeedID15 data set. consisting of images of cottonweed plants. For extracting fea-tures, they used methods like Image Segmentation, Histogram of Oriented Gradients (HOG), and gradient-weighted class activation. They were classified was performed by a self made ANN and several DCNN models for comparison. DCNN model Xception re-sulted in the best performance with a test set accuracy of 98.76 percent. The DCNN model could correctly classify the weeds 9 out of 10 times. The authors suggested exploring re-current neural networks and transfer learning models in the future.

Wang, Yecheng, etal. worked on one of the latest feature extractors, WinTransformer. Their survey revealed CNNs to be absolute for weed classification tasks, and transformers are the future. They used 2 datasets: 1) Self-collected maize/weed field images from China; 2) The Plant Seedlings dataset. In the initial phase, they fine-tuned the WinTransformer net-work using the Plant Seedlings dataset to create a task-specific pre-trained network. In the subsequent phase, they further fine-tune this pre-trained network on the MWFI dataset while incorporating a contrastive loss to aid in network training. To improve weed recogni-tion accuracy and minimize training data needs, they introduced a two-stage transfer learn-ing approach. Initially, they utilized a Swin Transformer network, pre-trained on ImageNet as input for the transfer learning stages. Experimental results displayed recognition accuracy, precision, recall, and an F1 score of 99.18, 99.33, 99.11, and 99.22 percentages, respec-tively. Yu and team aimed for a more robust and real-time weed classification. They sur-veyed earlier versions of YOLO and used the latest YOLOv5 model that had been released. They explained the importance of the multi-head self-attention mechanism (MHSA) and applied it to the Yolo backbone. The attention mechanism is the backbone of Yolov5. MHSA divides the input feature map into multiple heads. Each head calculates a different attention matrix, for capturing the relationship between different positions. CottonWee-dID15 data set has been used in this experiment. Although the peak accuracy is fairly low, compared to other approaches, experimental comparisons between the original YOLO v5 and improved versions show an increase of 1 percent.

Rizvi, Syed Mujtaba Hassan, et al. experimented with machine learning classifiers with manual and deep learning-based feature extraction methods. The preprocessing involved the annotation of CottonWeedID15 dataset, gray scaling, background removal using the U2Net model, and finally SMOTE for class balancing. Manual features like HuMoments and GLCM were extracted, and ANN showed superior performance compared to other classifiers. A transfer learning approach was used for automated feature extraction. ConvNeXt with RF as a classifier outperformed both manual and other DL models with a testing accuracy of 98 percent. The YOLOv8-M model was also trained, which showed 89 average precision. Graphical insights reveal that the removal of background showed no surge in accuracy but rather a decline in some cases.

Srivastava, Dhiraj, etal. compared 3 distinct vision approaches: ViT (Vision Transformers), MLP-mixers (Image Classification Networks), Yolo (Object Detection Model), and SimCLR, which is a self- supervised deep-learning technique. A dataset of 19,000 images was produced in fields of Virginia for soyabean. Images were annotated as bounding boxes for the Yolo model. ViT and MLP mixer models were fed segmented patches and were trained from scratch. To extract image features, a pre-trained ResNet-18 network served as the backbone for SimCLR. Results display that ViT outperformed MLP-mixer with testing accuracy of 97.95 percent and being 2.8 times inference speed. The YOLOv6s model achieved an AP of 0.811 for common ragweed detection.

Researchers have predominantly focused on detecting weeds or classifying them based on physical leaf characteristics, but the critical aspect of species classification has been overlooked. The species-level classification could significantly enhance precision agriculture. Species-level classification provides insight into weed ecology and distribution patterns, aiding in the development of predictive models for weed management. By refining our ability to identify weed species accurately, we can improve agricultural productivity and environmental stewardship.

Another observation is that in contemporary weed detection and classification research, Convolutional Neural Networks (CNNs) have been the primary focus, overshadowing the potential of recent transformer architectures.

3 Proposed Methodology

Based on the observations and identified research gap, we are proposing an updated DETR model.

3.1 Architecture

The objective is to use Detection transformer (DETR)-based model for weed identification. DETR architecture consists of following blocks: Convolutional Neural Network Block for feature extraction, transformer encoder-decoder, and feed-forward neural network for final detection. (Nicolas Carion et al.) The approach considers object detection as a problem of directly set prediction, eliminating the need for hand-crafted features such as anchor generation or non-maximum suppression procedure. Bipartite matching loss is used in DETR to measure the accuracy of the predicted boxes. It combines the box regression and classification loss and is computed using Hungarian algorithm.

However, in the proposed system, Vision Transformer has been used as a backbone instead of a Convolution Neural Network (CNN) block in DETR . The Vision Transformer uses patch embedding which takes an image of size H $\times$ W and divides it into N = H$\times$W /P2 non-overlapping patches, where P is the patch size. Each patch is then flattened into a vector of dimension D, and a learnable linear projection is applied for mapping the patch vector to a higher-dimensional space with E dimension (Dosovitskiy et al.).

Vision transformer uses an attention mechanism (Fig.2) that is based on the notion of self-attention, which enables the model to compute a weighted sum of the image

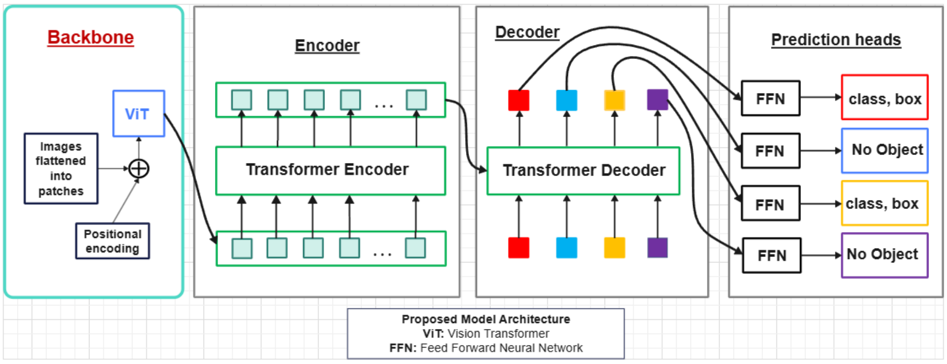

Fig. 1 Architecture of the proposed system. Here, ViT is used instead of CNN DETR backbone block

features at each position, where the weights are learned adaptively based on the input image. The attention mechanism computes a set of attention weights for each position in the image feature map, which determines how much attention should be paid to that position when making a prediction. These attention weights are then used to weigh the features at each position and compute a weighted sum, which is used as input to subsequent layers in the network.

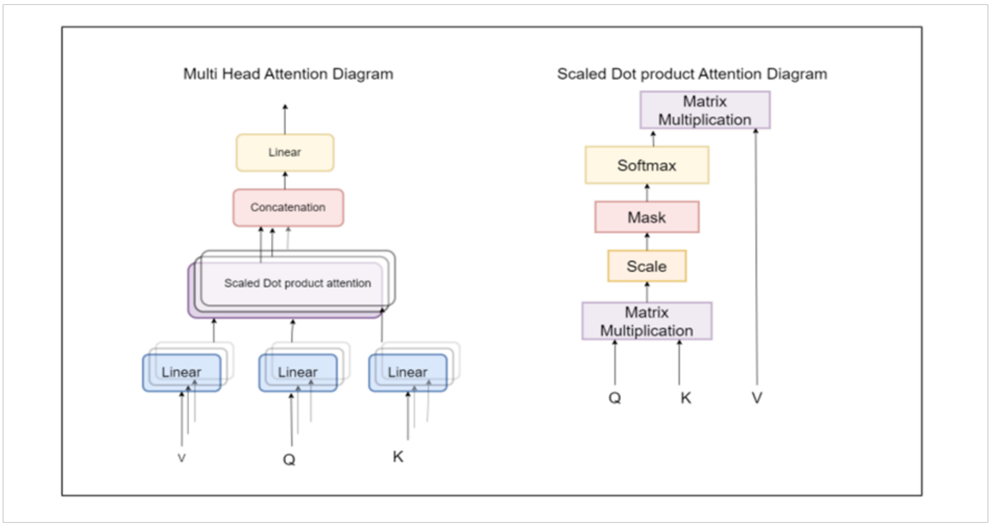

Fig. 2 Attention Mechanism used by ViT to weigh the features at each position and compute a weighted sum

The proposed solution will be evaluated against already existing state of art models in object detection considering parameters such as speed, accuracy, and evaluation metrics.

3.2 Experimental Setup

The study utilizes multiple publicly available datasets for training and evaluating weed classification and detection models. These datasets encompass a diverse range of weed species across different crop environments, providing a robust foundation for deep learning-based weed identification. The details of the datasets, including the number of images and class distributions, are summarized in Table I.

Table 1 Datasets used for weed classification and detection

Dataset Name	Dataset Size	Number of Weed Images
Soybean and weed Dataset. (Santos et al.)	400	Total: 15336, Broadleaf: 1191, Grass: 3520, Soybean: 7376, Soil: 3249
DeepWeeds dataset (Alex Olsen et al.)	17509	Total: 17509 (with 8 weed species)
Lincoln beet dataset (Gomez et al.)	4402	Total: 4402 (of sugar beets and weeds; labels in COCO JSON format)
Lettuce, Corn and Radish weed dataset (Honghua Jiang et al.)	7200	Corn: 6000 Lettuce: 800 Radish: 400

3.3 Classification and Detection Models

Latest versions of ViT-B32 and ViT-S16 models were utilized, along with the AlexNet model, which was developed using Keras and Tensorflow frameworks in Python. Vit-B32 and Vit-S16 models were loaded from TensorFlow Hub as feature extractors with pre-trained ImageNet weights for fine-tuning purposes. To fairly compare the performances of the models, all three models were trained with identical hyperparameters.

- 5-fold stratified cross-validation
- epochs per fold: 15
- batch size: 8
- Loss Function

The DETR model implementation of facebook research was used along with the proposed model, which is developed on top of DETR. Both models were loaded with pre-trained Im-ageNet weight for a backbone for fine-tuning. To ensure a fair

Table 2 Validation accuracy for each fold of three models

Fold No	Alex Net Model	ViT-S16	ViT-B32
1	93.44 ±3.03	98.92 ±1.15	99.57 ±0.17
2	94.089 ±1.98	98.77 ±1.15	99.59 ±0.15
3	94.69 ±2.13	98.92 ±0.79	99.46 ±0.46
4	93.46 ±2.69	99.01 ±0.79	99.52 ±0.07
5	94.41 ±0.81	99.00 ±0.788	99.82 ±0.32

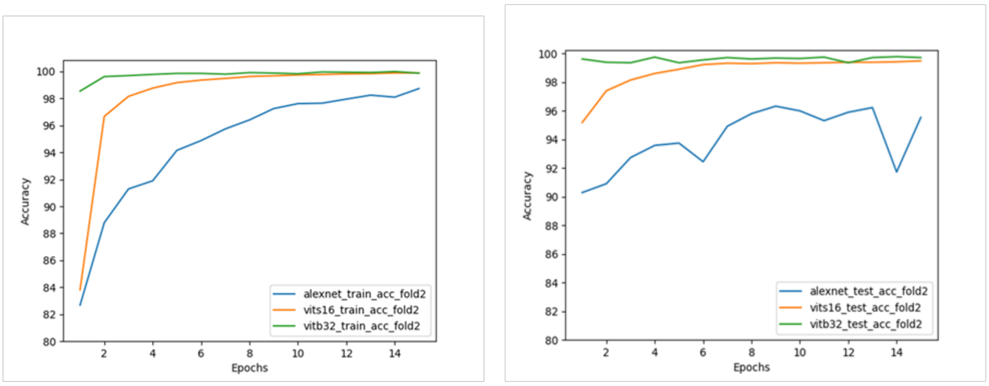

Fig. 3 Training vs testing accuracy of the Classification models

comparison of the model per-formances, models with identical hyperparameters were trained and their results were compared with models released by original paper (Abdur Rahman et.al).

- epochs: 100
- batch size: 8
- Loss Function

4 Results

4.1 Performance evaluation of Classification Models

The evaluation involved training three different models using a five-fold cross-validation. The models were trained on the training set of each fold and evaluated on the corresponding testing set. Visualizations of the training and testing accuracy comparisons for each model across the folds were plotted. Figure.3 shows the graphs that display the training and testing performance comparison of three models for four folds of cross-validation. Also, testing accuracies for each fold for 3 models and the confusion matrix are given below. From the results, it can be concluded that ViT-B32 achieved an overall accuracy of 99.93 percent, followed by ViT-s16 with an accuracy of 99.78 percent and AlexNet with 98.72 percent accuracy.

4.2 Performance evaluation of Detection Models

The evaluation involved training two models, namely the DETR and the proposed model using cross-validation. Then, plotted bounding box predictions by a proposed model with respect to ground truth were plotted. Class- wise AP scores and mAp for models and plotted AP score, precision, recall, F1 score, and loss function for the proposed model were also calculated (Fig.4). From the results, it can be concluded that the proposed model ViT-DETR achieved an mAP (IoU: 0.5) of 82.1 percent followed by DETR (mAP (IoU: 0.5): 79 percent).

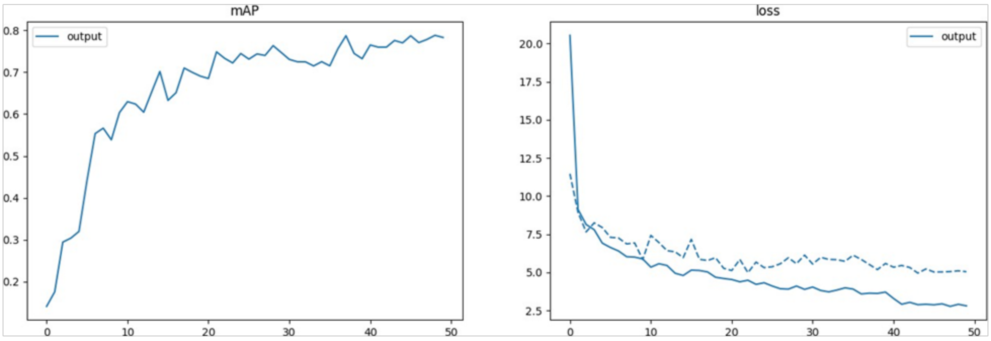

Fig. 4 Mean Average Precision (mAP) progression over training epochs (left) and the loss function convergence (right) for the proposed ViT-DETR model

5 Conclusion

Advanced deep learning techniques have shown promising results for weed identification. Conventional approaches face significant challenges like not having enough data and struggling to distinguish between weeds and crops due to similar properties. To tackle this, a fusion of vision transformer and detection transformer is used, DETR for accurate localization of weeds and Vision Transformer for identifying different types of weeds. The proposed method improved the accuracy on the test dataset to 82.9 percent. This beats the performance of traditional CNN approaches, which shows a big step forward in making weed identification more accurate for sustainable agriculture practices.

References

[1] A. Ahmad, D. Saraswat, V. Aggarwal, A. Etienne, and B. Hancock, "Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems," *Computers and Electronics in Agriculture*, vol. 184, p. 106081, 2021.

[2] A. Borthakur and P. Singh, "Agricultural research in India: An exploratory study," *International Journal of Social Science and Interdisciplinary Research*, vol. 1, no. 9, pp. 59–74, 2012.

[3] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, UK, 2020, pp. 213–229.

[4] D. Chen, Y. Lu, Z. Li, and S. Young, "Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems," *Computers and Electronics in Agriculture*, vol. 198, p. 107091, 2022.

[5] A. dos Santos Ferreira, D. M. Freitas, G. G. da Silva, H. Pistori, and M. T. Folhes, "Weed detection in soybean crops using convnets," *Computers and Electronics in Agriculture*, vol. 143, pp. 314–324, 2017.

[6] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[7] Y. Gharde, P. K. Singh, R. P. Dubey, and P. K. Gupta, "Assessment of yield and economic losses in agriculture due to weeds in India," *Crop Protection*, vol. 107, pp. 12–18, 2018.

[8] H. Jiang, C. Zhang, Y. Qiao, Z. Zhang, W. Zhang, and C. Song, "CNN feature-based graph convolutional network for weed and crop recognition in smart farming," *Computers and Electronics in Agriculture*, vol. 174, p. 105450, 2020.

[9] P. Lameski, E. Zdravevski, V. Trajkovik, and A. Kulakov, "Weed detection dataset with RGB images taken under variable light conditions," in *Proc. 9th Int. Conf. ICT Innov.*, Skopje, Macedonia, 2017, pp. 112–119.

[10] A. Olsen *et al.*, "DeepWeeds: A multiclass weed species image dataset for deep learning," *Scientific Reports*, vol. 9, no. 1, p. 2058, 2019.

[11] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?," in *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 12116–12128, 2021.

[12] A. Rahman, Y. Lu, and H. Wang, "Performance evaluation of deep learning object detectors for weed detection for cotton," *Smart Agric. Technol.*, vol. 3, p. 100126, 2023.

[13] N. Razfar, J. True, R. Bassiouny, V. Venkatesh, and R. Kashef, "Weed detection in soybean crops using custom lightweight deep learning models," *J. Agric. Food Res.*, vol. 8, p. 100308, 2022.

[14] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, “Transformer neural network for weed and crop classification of high-resolution UAV images,” *Remote Sens.*, vol. 14, no. 3, p. 592, 2022.

[15] A. Salazar-Gomez *et al.*, “Towards practical object detection for weed spraying in precision agriculture,” *arXiv preprint arXiv:2109.11048*, 2021.

[16] T. Sarvini *et al.*, “Performance comparison of weed detection algorithms,” in *Proc. 2019 Int. Conf. Commun. Signal Process. (ICCSP)*, 2019, pp. 843–847.

[17] A. K. Singh, “Precision agriculture in India – Opportunities and challenges,” *Indian J. Fertil.*, vol. 18, no. 4, pp. 308–331, 2022.

[18] S. Srivastava *et al.*, “Comparative analysis of deep learning image detection algorithms,” *J. Big Data*, vol. 8, no. 1, pp. 1–27, 2021.

[19] A. Vaswani *et al.*, “Attention is all you need,” in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[20] P. Wang *et al.*, “Weed25: A deep learning dataset for weed identification,” *Front. Plant Sci.*, vol. 13, 2022.

[21] Y. Wang, Y. Chen, and D. Wang, “Convolution network enlightened transformer for regional crop disease classification,” *Electronics*, vol. 11, no. 19, p. 3174, 2022.

[22] R. T. U. Mahmud *et al.*, “Current and future applications of statistical machine learning algorithms for agricultural machine vision systems,” *Comput. Electron. Agric.*, vol. 156, pp. 585–605, 2019.

[23] D. Patel and B. Kumbhar, “Weed and its management: A major threat to crop economy,” *J. Pharm. Sci. Bioscientific Res.*, vol. 6, no. 6, pp. 753–758, 2016.

[24] R. L. Zimdahl, *Fundamentals of Weed Science*, 5th ed. Academic Press, 2018.

[25] N. Iqbal *et al.*, “Investigation of alternate herbicides for effective weed management in glyphosate-tolerant cotton,” *Arch. Agron. Soil Sci.*, vol. 65, no. 13, pp. 1885–1899, 2019.

[26] R. Punithavathi *et al.*, “Computer vision and deep learning-enabled weed detection model for precision agriculture,” *Comput. Syst. Sci. Eng.*, vol. 44, no. 3, pp. 2759–2774, 2023.

[27] P. Ong, K. S. Teo, and C. K. Sia, “UAV-based weed detection in Chinese cabbage using deep learning,” *Smart Agric. Technol.*, vol. 4, p. 100181, 2023.

[28] N. P. *et al.*, “Classification of weed detection control management using artificial and deep convolutional neural networks,” in *Proc. 9th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Coimbatore, India, 2023, pp. 1294–1299.

[29] M. Sanjay *et al.*, "Weed detection: A vision transformer approach for soybean crops," in *Proc. 14th Int. Conf. Comput., Commun., Netw. Technol. (ICCCNT)*, 2023, pp. 1–8.

[30] S. M. H. Rizvi *et al.*, "Revolutionizing agriculture: Machine and deep learning solutions for enhanced crop quality and weed control," *IEEE Access*, 2024.

[31] P. Khaire, V. Attar, and S. Kalamkar, "A comprehensive survey of weed detection and classification datasets for precision agriculture," in *Proc. Int. Conf. Comput., Commun., Netw. Technol. (ICCCNT)*, 2023, pp. 1–5.

[32] K. Osorio *et al.*, "A deep learning approach for weed detection in lettuce crops using multispectral images," *AgriEngineering*, vol. 2, no. 3, pp. 471–488, 2020.

[33] M. H. Asad and A. Bais, "Weed detection in canola fields using maximum likelihood classification and deep convolution neural network," *Inf. Process. Agric.*, vol. 7, no. 4, pp. 535–545, 2020.