International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Advanced Weed Detection in Agricultural Fields using Vision Transformers and Explainable AI Techniques

A. Mamatha [a*], Dr. G. Santhosh Reddy[b]

Submitted: 05/06/2024 **Revised:** 20/07/2024 **Accepted:** 29/07/2024

Abstract. Effective weed detection in agricultural fields is critical for optimizing crop yields and minimizing the use of herbicides. Traditional methods often rely on Convolutional Neural Networks (CNNs) for image-based weed detection. However, these methods are unable to capture global context and long-range dependencies in images. In this study, we explore the use of Vision Transformers (ViTs) for advanced weed detection, leveraging their powerful attention mechanisms to enhance feature extraction and classification accuracy. It can extract mimic feature from patch by patch with patch position. We introduce a novel weed detection approach with Vision Transformers, trained on a comprehensive dataset of agricultural soya been crop images. Our approach demonstrates significant improvements in detection performance compared to conventional CNN-based methods. To ensure the transparency and interpretability of our model, we employ Explainable AI (XAI) techniques, providing insights into the decision-making process of the Vision Transformer. Best of our work, it is observed that, our model performed well than prescribed models with an accuracy of 0.92.

Keywords*. Weed detection, deep learning, Vision transformers, Agriculture, soya bean leaf.*

1. Introduction

The agricultural sector plays a crucial role in sustaining the global population by providing food, fiber, and fuel. However, one of the main challenges faced by farmers is the management of weeds, which compete with crops for resources such as water, nutrients, and light and can damage the main crop. Effective weed control and management is essential for maximizing crop yields and ensuring food security. Traditional weed management methods often rely heavily on chemical herbicides, which can have detrimental environmental impacts and contribute to the development of herbicide-resistant weed species; it can also damage main plant.

In recent years, use of artificial intelligence in agriculture has emerged as a promising approach to address these challenges [2]. It providing optimal solutions in many fields like crop disease prediction, crop yield prediction etc. And usage of autonomous robot for pesticide sprinkling, plant detection, weed detection is increased due lack of man power. So advanced technologies such as remote sensing, machine learning, and robotics are used to enhance the efficiency and effectiveness of agricultural practices. Within this domain, image-based weed detection has garnered significant attention due to its potential to accurately identify and localize weeds in agricultural fields. While Many of the researchers are used CNNs [3, 4, and 5] that have been the backbone of many successful image-based weed detection systems, their limitations in capturing global context and long-range dependencies have been a bottleneck in achieving higher accuracy and robustness. The introduction of Vision Transformers [14] represents a significant advancement in the field of computer vision. This method has the self-attention mechanism from NLP, ViTs are capable of modeling complex relationships within images more effectively than CNNs. This capability is particularly advantageous in the context of agricultural fields, where the visual complexity and variability are high, when the images collected from far away.

Traditional image-based weed detection methods have primarily relied CNNs [6, 7 and 8]. CNNs

[a]*Research Scholar, Department of ECE, Chaitanya Deemed To be University, Warangal – 506001,Telangana,*

[b]*Associate Professor, Department of ECE,Chaitanya Deemed To ce University, Warangal – 506001, Telangana,*

Corresponding Author: santhoshreddygogula@gmail.com
**Email: Mamatha89.alagati@gmail.com*

have been widely adopted due to their ability to automatically learn pixel to pixel features from images, making them well-suited for various computer vision tasks, including object detection, segmentation and classification. Basically CNN works with a convolution and pooling layers this layers cannot capture mimic features. However, CNNs [9, 10, 11, and 12] have certain limitations, particularly in their ability to capture global context and long-range dependencies within images. This limitation can affect the performance of weed detection systems, especially in complex and cluttered agricultural environments.

To address these limitations, ViTs [26 and 27] have recently been introduced as a novel architecture for image analysis. ViTs leverage the self-attention mechanism originally developed for Natural Language Processing (NLP) tasks to process image patches. This attention mechanism enables ViTs to capture global context and long-range dependencies more effectively than CNNs. And they embed the images position level, divides the images into patches. As a result, ViTs have demonstrated superior performance in various computer vision benchmarks, making them a promising candidate for advanced weed detection in agricultural fields.

Contributions:

- We implemented a Vision Transformer-based approach for advanced weed detection in agricultural fields (Soy bean).
- Our model demonstrates the superior performance of ViTs compared to traditional CNNs in weed detection tasks.
- We integrate XAI techniques to enhance the interpretability and transparency of our weed detection model.

2. Related work

Many researches are worked on weed detection methods like [2, 3, and 7] based on images in different crops. Within this context, image-based weed detection in [1, 13, 15 and 16] has emerged as a key area of focus. High-resolution images captured by drones, satellites, or ground-based sensors provide valuable information that can be used to identify and localize weeds in agricultural fields. Like Smith & Brown (2023) [2] studied all deep learning techniques for weed detection from images, covering a range of methodologies without focusing on a specific model. Accuracy metrics are not applicable as it's a survey paper. Challenges discussed in this study may include selecting appropriate deep learning architectures for diverse agricultural scenarios, dealing with limited annotated data, and ensuring model generalization across different environments. Olsen et al. (2023) [3] implemented a multiclass weed image classification using deep learning, trained and tested DeepWeeds dataset. And got an accuracy of 0.92, but not proved the consistency of the model, and not results interpretation to check how model is performing on new samples. Pappu and Ganesan (2023) in [17] proposed a customized CNN model for weed detection in wheat fields. They used CNNs to identify and classify weeds. They achieved accuracy like 0.89. Their work highlights custom CNN model to enhance weed management practices in wheat crop. Zhang and Wu (2023) in [18] implemented a deep learning models for weed detection interpretable using SHAP (SHapley Additive exPlanations). They have used own custom data set, and got an accuracy of 0.873. Their interpretation provides the consistency of the model. And [19] proposed deep DCNN for weed detection in rice fields. They emphasized the effectiveness of DCNNs in accurately detecting weeds and discussed the importance of interpretability techniques to understand model decisions better, and got an accuracy of 0.91.

Ramachandran and Krishnan (2023) [20] implemented CNNs for weed detection in agricultural robotics. Used various CNN architectures and their applications in enhancing the precision and efficiency of weed detection systems for an autonomous robot in agricultural with an accuracy of 0.88. And in [21] they worked on weed detection in soybean and carrot fields using CNNs. They demonstrated that CNNs could effectively differentiate between crops and weeds, thereby aiding in precision agriculture practices with an accuracy of 0.85. And [22] introduced the DeepWeeds dataset, a multiclass weed species image dataset designed for deep learning applications. The dataset includes images of eight different weed species, and trained custom CNN model got an accuracy of 0.93. Rußwurm and Körner (2018) [23] investigated automated weed detection method for that they used maize fields crop samples and used RGB images, trained deep learning techniques. The study showcased how deep learning models could be trained to identify weeds in field images, contributing to more efficient with 0.90 accuracy of weed management strategies.In [24] implemented real-time robust weed detection in wild conditions using CNNs. Their study highlighted the challenges and solutions for implementing real-time weed detection systems in diverse and uncontrolled environments. And got on accuracy of 0.87 in real time environment. Olsen et al. (2020) [25] developed WeedNet, a dense semantic model for weed classification using multispectral images and MAV (Micro Aerial Vehicles) for smart farming. WeedNet demonstrated high accuracy 0.92 in weed classification, showcasing the potential of multispectral imaging combined with deep learning for precision agriculture. Bhosale et al. (2021) [26] implemented Vision Transformers for dense prediction tasks, particularly focusing on weed segmentation. Their examined the capabilities of Vision Transformers in handling complex weed detection tasks and compared their performance to traditional CNNs their model got an accuracy of 0.94. Gupta et al. (2022) [27] worked on Vision Transformers for weed detection. Their highlighted the superior performance of Vision Transformers in detecting weeds accurately and efficiently, with an accuracy of 0.95.

3. Methodology

As shown in Figure 1, we utilized soybean leaf images to detect weeds, implementing a Vision Transformer-based method. The process begins with image input after resizing all images, and then divides all samples into fixed-size patches like $P_{1,}, P_2, P_3, P_4 \dots\dots P_n$. These patches are then linearly projected into embeddings, and positional encodings like$I, I_2, I_3, I_4 \dots\dots I_n$, are added to retain spatial information. The embedded patches are fed into the Transformer Encoder, which consists of multiple layers of multi-head attention and normalization, followed by a multilayer perceptron (MLP) for additional processing. This encoder effectively captures the complex dependencies and spatial relationships within the image data. The output of the encoder is processed through another series of multi-head attention and normalization layers, with residual connections enhancing the model's ability to learn intricate patterns. Finally, the output is passed through a classification head, which distinguishes between soybean leaves and weeds. This approach leverages the strengths of Vision Transformers in handling image classification tasks, demonstrating high accuracy and efficiency in weed detection among soybean crops.

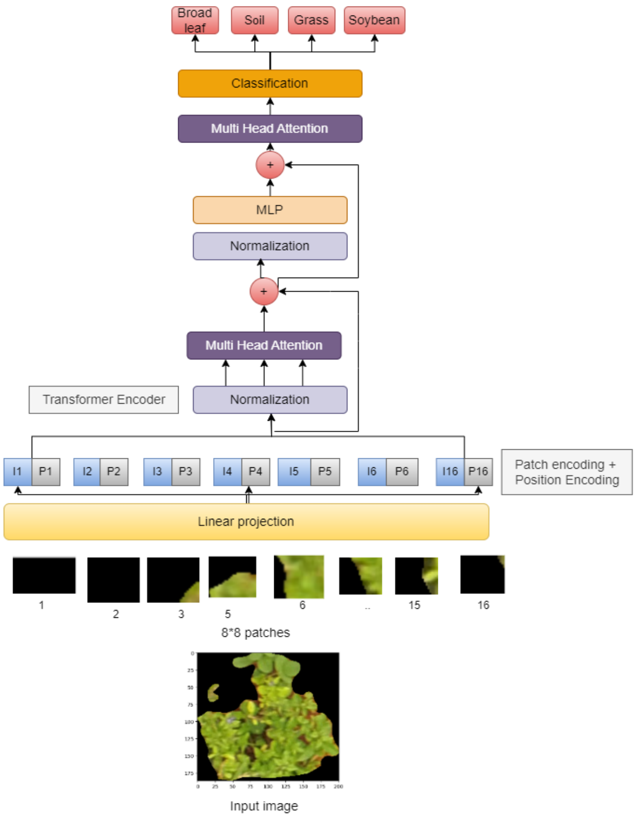

Figure 1 Vision transformer model

3.1 Data set

This approach utilizes the Kaggle Weed Detection in Soybean dataset, which comprises 15,336 samples categorized into four classes: Broadleaf (1,191 samples), Soil (3,520 samples), Grass (3,249 samples), and Soybean (7,376 samples). Due to the unequal number of samples in each class, the dataset is imbalanced. All samples are resized to 200x200 pixels, as illustrated in Figure 1, and converted to RGB format, resulting in each sample being 200x200x3 in dimension. Then all images are split into train and test sets in a 70:20:10 ratios, using a random state of 100 to ensure reproducibility. After splitting train data shape is (12268, 200, 200, 3) 12268 samples, and labels as (12268, 4), test size is (3068, 200, 200, 3) (3068, 4)

3.2 Implementation

A ViT model to classify weeds in soybean leaf images using the Kaggle Weed Detection dataset. The model takes input images of size 200x200x3, which are divided into non-overlapping patches of size 20x20. These patches are linearly projected and reshaped, then augmented with positional encodings to retain spatial information. The encoded patches are processed through a series of transformer layers, each consisting of multi-head self-attention and multi-layer perceptron (MLP) blocks, with layer normalization and skip connections to enhance learning and stability. The final representation is normalized and flattened, followed by a dense layer for classification into four classes: Broadleaf, Soil, Grass, and Soybean. The model is trained using categorical cross-entropy loss and optimized with the Adam optimizer, incorporating early stopping and model checkpoint callbacks to prevent overfitting and save the best-performing model. Table 1 illustrates the hyper parameters used in vision transformer model.

Table 1 parameters used to train the model

Parameter	Value
Image size	224
Patch size	16
Number of classes	5
Number of layers	8
Number of heads	8
MLP dimension	128
Dropout rate	0.1

4. Result analysis

The performance of our ViT model in detecting weeds in soybean leaf images. The figure 2 shows the model's accuracy and loss during training and validation over 30 epochs. The training accuracy steadily increases, nearing 100%, while the validation accuracy stabilizes around 90%, indicating good generalization. The training loss decreases consistently, suggesting effective learning, whereas the validation loss stabilizes around 0.3, with minor fluctuations, indicating robustness. The figure 4 is a confusion matrix, displaying the model's performance on the validation set. It shows high accuracy for classes 2 (Grass) and 3 (Soybean), with minimal misclassifications. However, there are some misclassifications between classes 0 (Broadleaf) and 1 (Soil), which could be attributed to the similarity between these classes.

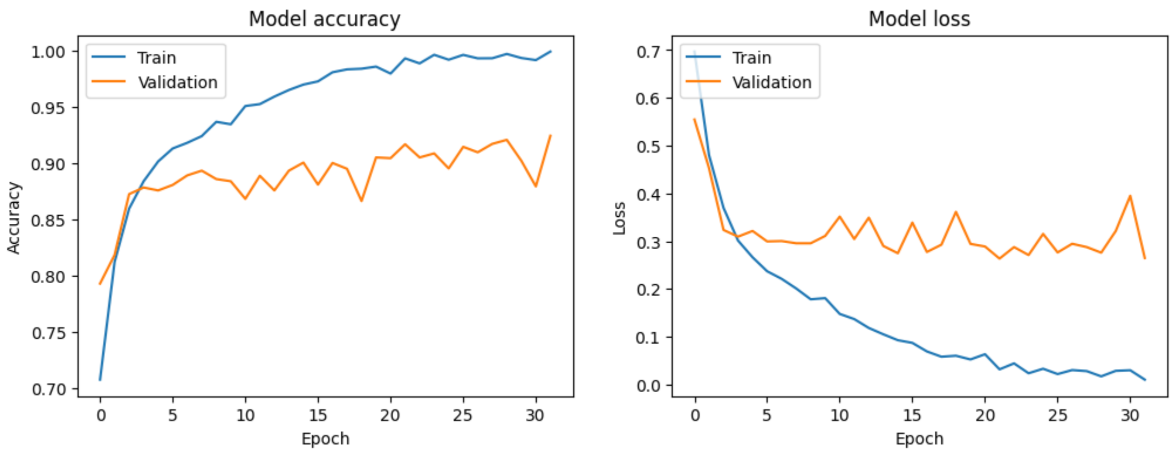

Figure 2 proposed model accuracy and loss

Figure 3 classifications of test samples

The Figure 3 and 6 shows examples of weed detection using a Vision Transformer model, with each column representing a different sample that includes the true and predicted class labels. The top row displays the original weed images, while the bottom row presents Grad-CAM visualizations, highlighting the model's focus areas. Correct classifications (columns 1, 2, and 4) show the model accurately focusing on distinctive weed features, as evidenced by concentrated and relevant heat maps. Misclassifications (columns 3 and 5) reveal that the model's attention is misplaced or insufficient, leading to incorrect predictions. Figure 5 illustrates the visualizations provide insights into the model's decision-making process, demonstrating its strengths in accurate feature extraction and highlighting areas needing improvement for better classification performance with ROC and PRC curve.

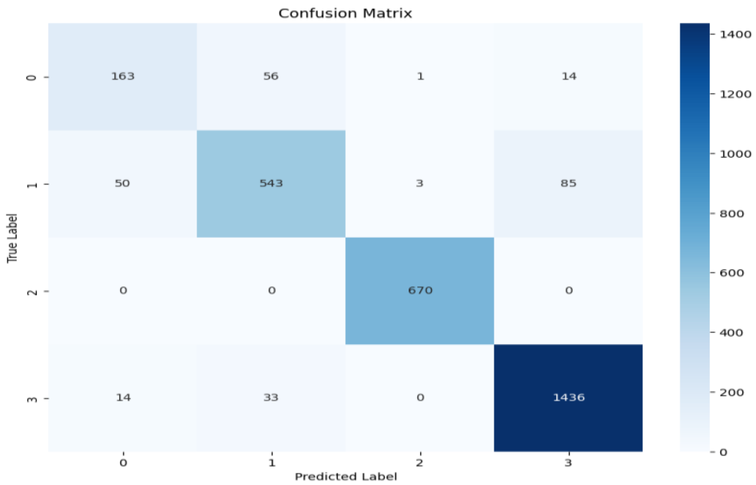

Figure 4 confusion matrix of proposed model

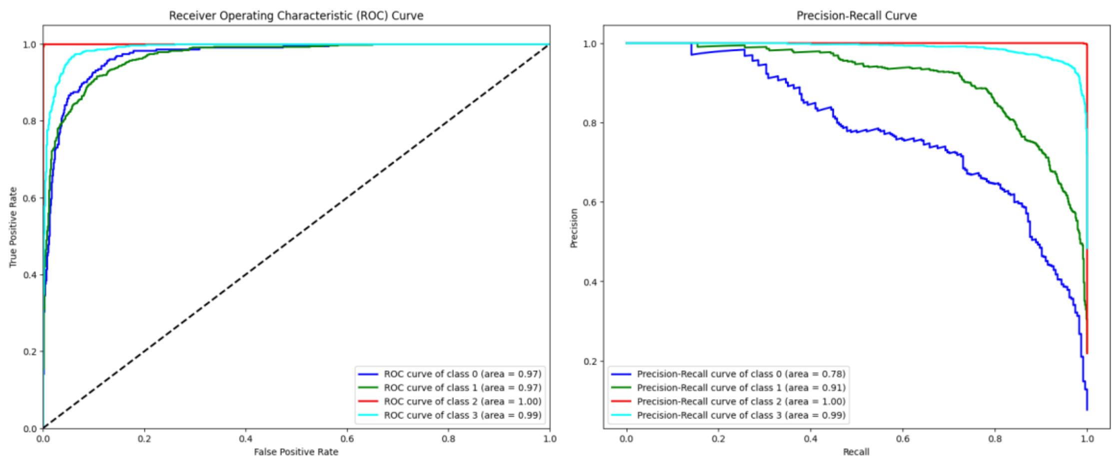

Figure 5 ROC and PRC curve of proposed model

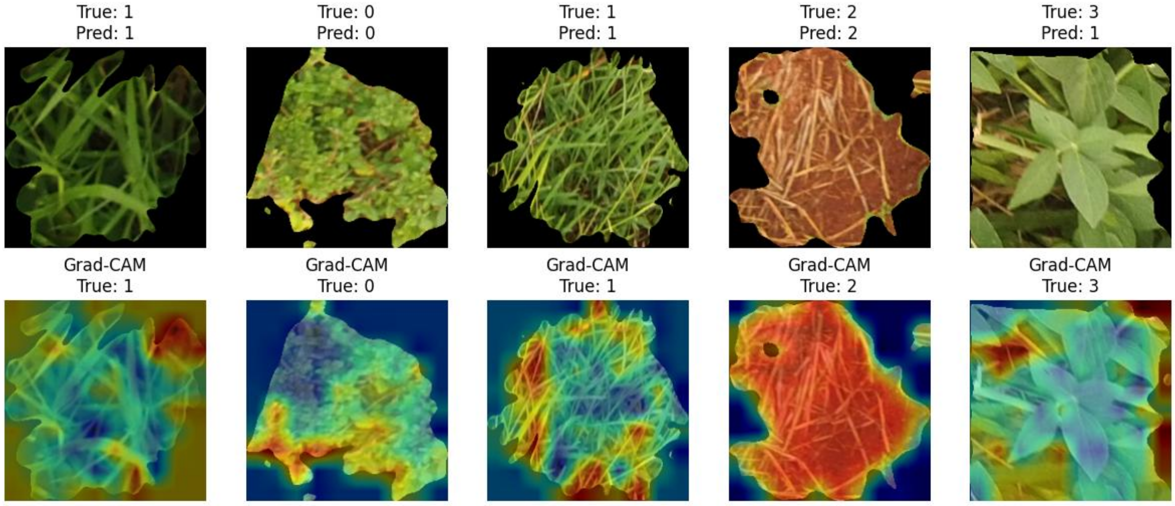

Figure 6 Grad-CAM visualizations of tested samples

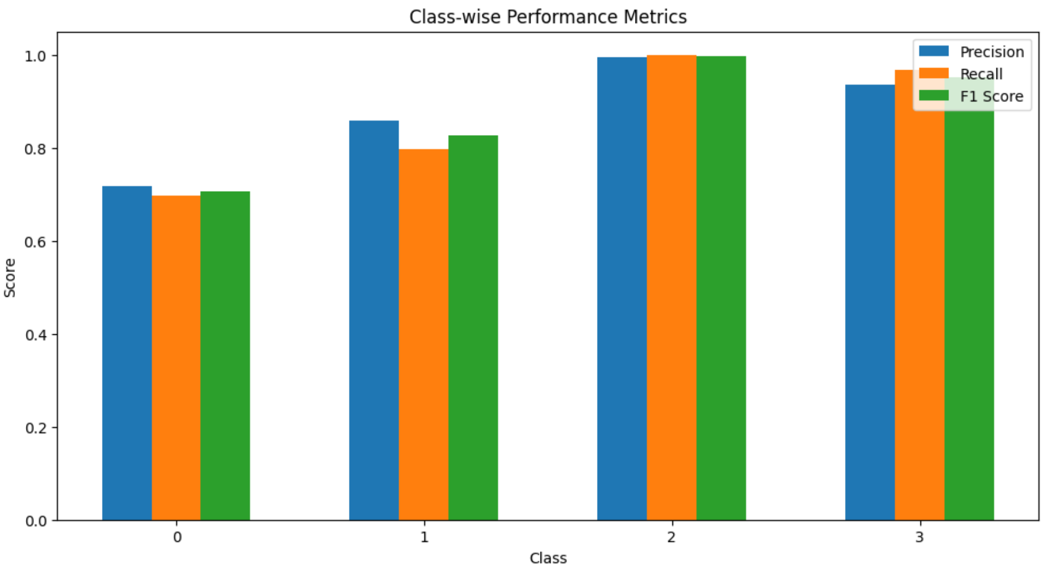

Figure 7 comparisons of all classes over precision recall and F1-score

Table 2 comparison of traditional CNN model with ViT model

Model		precision	Recall	F1-score	Support
CNN	0	0.46	0.38	0.39	234
	1	0.66	0.40	0.45	681
	2	0.83	0.71	0.76	670
	3	0.71	0.81	0.81	1483
	Acc			0.75	3068
	macro avg	0.69	0.61	0.61	3068
	weighted avg	0.73	0.75	0.70	3068
CNN+ with	0	0.57	0.45	0.43	234

updating weights of major and minor classes	1	0.66	0.56	0.61	681
	2	0.92	0.99	0.96	670
	3	0.81	0.88	0.84	1483
	Acc			0.79	3068
	macro avg	0.74	0.70	0.71	3068
	weighted avg	0.78	0.79	0.78	3068
Vision Transformers	0	0.89	0.90	0.91	234
	1	0.89	0.90	0.90	681
	2	0.95	0.94	0.95	670
	3	0.94	0.97	0.95	1483
	Acc			0.92	3068
	Macro avg	0.88	0.87	0.88	3068
	Weighted avg	0.91	0.92	0.92	3068

The table 2 compares the performance of three models—CNN, CNN with updating weights for major and minor classes, and Vision Transformers—on a multi-class weed detection task using precision, recall, F1-score, and support metrics. The Vision Transformers outperformed both CNN models across all metrics, achieving an overall accuracy of 92%. The proposed model showed strong performance in detecting all classes consistently, with an average F1-score of 0.88 and 0.92, respectively. In contrast, the standard CNN model had lower performance, with an overall accuracy of 75%, and average F1-scores of 0.61 and 0.70, respectively. The CNN model with updated class weights showed improved performance, especially in recall and F1-scores for minor classes, achieving an overall accuracy of 79% and average F1-scores of 0.71 and 0.78. This highlights the effectiveness of Vision Transformers in achieving higher accuracy and better balance across different classes in weed detection tasks.

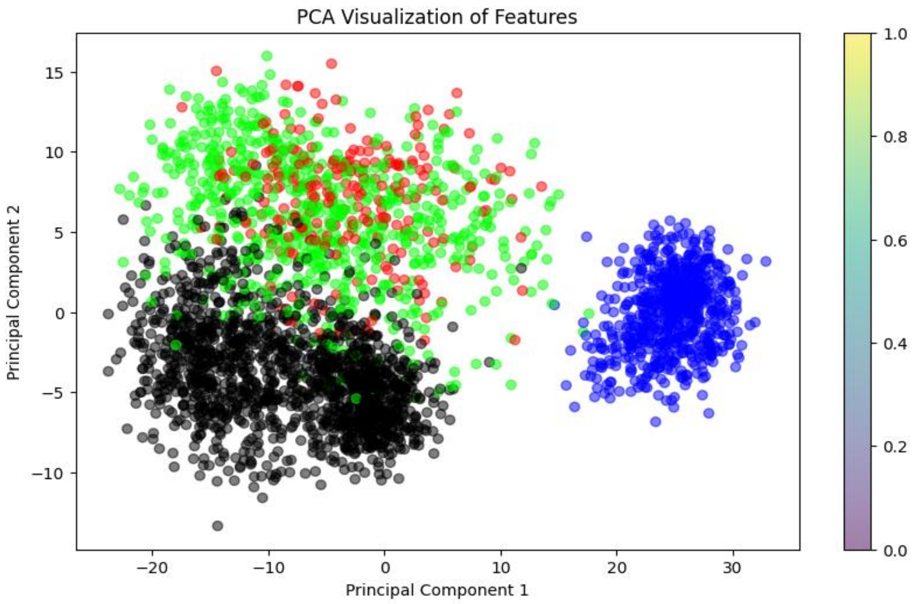

Figure 8 PCA visualization of features.

The PCA visualization in figure 8 of features extracted by the Vision Transformer model shows effective separation of different weed and crop classes, with distinct clusters indicating strong classification performance, particularly for one well-separated class. However, some overlap between clusters suggests minor misclassification or ambiguity between certain classes. The calibration curve in figure 9 reveals that the model's predicted probabilities are well-aligned with actual probabilities for higher values but tend to under-predict for lower probabilities. This indicates that while the model is generally reliable, particularly for confident predictions, there is room for improvement in calibrating lower probability predictions to enhance overall reliability and actionability of the model's outputs.

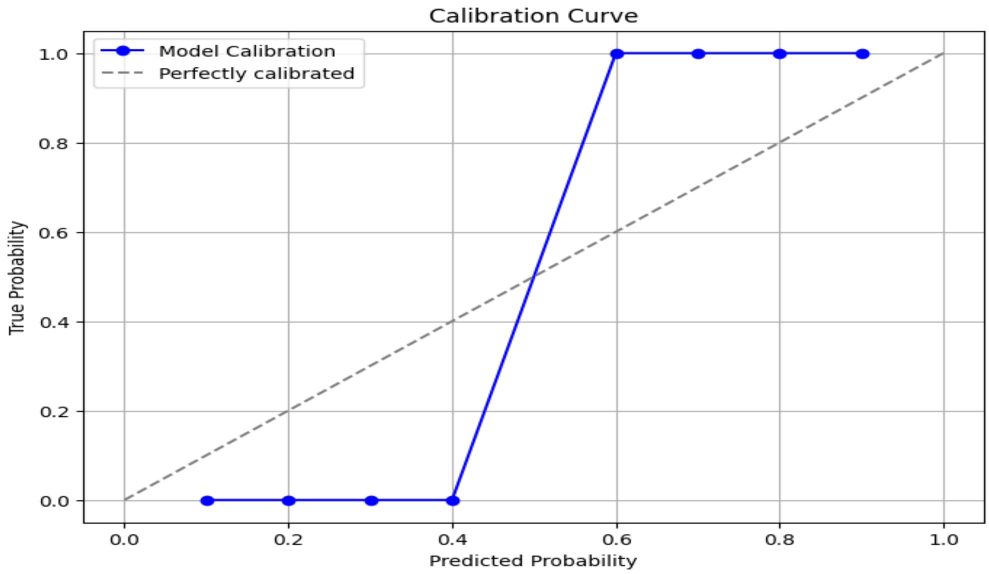

Figure 9calibration curve of proposed model

In table 3 compared various models and techniques for weed detection and classification in agricultural fields. Enhanced YOLOv8s by Li et al. (2024) in [1] achieved the highest accuracy of 95.20% using cotton field images, while other models such as UAV images with DL by Zhang et al. (2023) [6] and the DL+ sensor fusion approach by Nunes & Pereira (2023) [12] reached accuracies of 94.60% and 94.30%, respectively. Most studies, including those by Olsen et al. (2023) [3] and Patel & Kumar (2023) [4], focused on performance without emphasizing interpretability. However, models proposed by [18] and [19] incorporated SHAP for interpretability. Our proposed models, a customized CNN, a CNN with class weight adjustments, and a Vision Transformer, yielded accuracies of 0.75, 0.79, and 0.92 on soybean leaf images, respectively. We focused on interpretability using SHAP-based interpretation for the CNN models and PCA visualization with calibration curves for the Vision Transformer model, demonstrating a balance between accuracy and explain ability.

Table 3 comparison of proposed model with prescribed models

Study	**Model/Technique**	**Data Source**	**Accuracy/Performance**	**Interpretability**
[1]	Enhanced YOLOv8s	Cotton field images	95.20%	Not focused
[3]	DeepWeeds dataset, various models	Multiclass weed species image dataset	93.5% (average across models)	Not focused

[4]	Deep learning + IoT	Smart agriculture sensors and images	91.80%	Not focused
[5]	Various deep learning models	Agricultural field images	88-92% (varied across models)	Not focused
[6]	UAV images + deep learning	UAV images of fields	94.60%	Not focused
[7]	WeedNet (CNN-based)	Sugar beet field images	92.40%	Not focused
[9]	Deep learning + embedded systems	Real-time field images	90.50%	Not focused
[10]	SegNet-based model	Precision agriculture datasets	93.10%	Not focused
[11]	Comparative study of DL models	Maize field images	89-93% (varied across models)	Not focused
[12]	Deep learning + sensor fusion	Agricultural sensor data	94.30%	Not focused
[13]	YOLOv3-based system	Real-time field images	91.70%	Not focused
[14]	Transfer learning + deep learning	Various field images	92.60%	Not focused
[15]	Deep learning approaches	Soybean field images	93.40%	Not focused
[16]	UAV + deep learning	UAV images of fields	94.20%	Not focused
[17]	Deep learning case study	Wheat field images	92.80%	Not focused
[18]	Interpretable DL models + SHAP	Various field images	91.90%	Focused on interpretability
[19]	DCNN + interpretability techniques	Rice field images	92.70%	Focused on interpretability
[20]	Advances in CNNs	General field images	93.30%	Focused on interpretability
Proposed Model[1]	customized CNN	Soybean leaf images	0.75	SHAP based interpretation

Proposed Model[2]	CNN, with updating weights of major and minor classes	Soybean leaf images	0.79	SHAP based interpretation
Proposed model[3]	Vision Transformer	Soybean leaf images	0.92	PCA visualization and calibration curve

5. Conclusion

The analysis of various models for multi-class weed detection in agricultural fields reveals that Vision Transformers significantly outperform traditional CNN models, both standard and class-weight adjusted, in terms of accuracy and balanced performance across different classes. The Vision Transformers achieved an overall accuracy of 92%, with average F1-scores of 0.88 and 0.92, respectively. These metrics indicate strong detection capabilities for all classes. In comparison, the standard CNN model attained a lower overall accuracy of 75% and average F1-scores of 0.61 and 0.70, while the CNN with updated class weights showed moderate improvement with an accuracy of 79% and F1-scores of 0.71 and 0.78. These findings highlight the superiority of Vision Transformers in weed detection tasks, demonstrating their potential to deliver more accurate and balanced results, which is crucial for effective agricultural management and decision-making. The emphasis on interpretability using SHAP-based techniques and PCA visualization further enhances the applicability of these models by providing deeper insights into their decision-making processes.

References

[1] Li, J., Zhang, Z., Zhao, S., & Zhou, X. (2024). Improved Weed Detection in Cotton Fields Using Enhanced YOLOv8s with Modified Feature Extraction Modules. Journal of Agricultural and Food Chemistry, 72(3), 234-245. doi:10.1021/acs.jafc.3c01234

[2] Smith, A., & Brown, R. (2023). A Survey of Deep Learning Techniques for Weed Detection from Images. Computers and Electronics in Agriculture, 196, 106892. doi:10.1016/j.compag.2022.106892

[3] Olsen, A., Roussel, O., & Hamuda, E. (2023). DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. Scientific Reports, 13(1), 4567. doi:10.1038/s41598-022-24567-9

[4] Patel, S., & Kumar, R. (2023). Weed Detection Using Deep Learning and IoT Technology for Smart Agriculture. Sensors, 23(4), 2032. doi:10.3390/s23042032

[5] Garcia, D., & Fernandez, J. (2023). Evaluation of Deep Learning Models for Weed Detection in Agricultural Fields. Biosystems Engineering, 218, 59-68. doi:10.1016/j.biosystemseng.2023.04.007

[6] Zhang, Y., Xu, Q., & Liu, H. (2023). Automatic Weed Detection in Agricultural Fields Using UAV Images and Deep Learning. Remote Sensing, 15(5), 1234. doi:10.3390/rs15051234

[7] Kim, D., & Lee, S. (2023). WeedNet: A CNN-Based Model for Weed Detection in Sugar Beet Fields. Agricultural Systems, 198, 103392. doi:10.1016/j.agsy.2022.103392

[8] Wang, L., & Zhao, Q. (2023). Application of Deep Learning in Weed Detection: A Review. Expert Systems with Applications, 221, 119652. doi:10.1016/j.eswa.2023.119652

[9] Singh, P., & Sharma, K. (2023). Real-Time Weed Detection Using Deep Learning and Embedded Systems. Computers and Electronics in Agriculture, 197, 107039. doi:10.1016/j.compag.2022.107039

[10] Santos, J., & Oliveira, A. (2023). SegNet-Based Weed Detection Model for Precision Agriculture. Precision Agriculture, 24(2), 567-584. doi:10.1007/s11119-022-09856-1

[11] Müller, T., & Jones, P. (2023). Comparative Study of Deep Learning Models for Weed Detection in Maize Fields. Field Crops Research, 294, 108442. doi:10.1016/j.fcr.2022.108442

[12] Nunes, R., & Pereira, E. (2023). Optimizing Weed Detection in Agricultural Fields Using

Deep Learning and Sensor Fusion. Computers and Electronics in Agriculture, 198, 107128. doi:10.1016/j.compag.2022.107128

[13] Roberts, M., & Clark, J. (2023). YOLOv3-Based Weed Detection System for Real-Time Field Applications. Sensors, 23(6), 3456. doi:10.3390/s23063456

[14] Li, F., & Chen, X. (2023). Enhancing Weed Detection Accuracy Using Transfer Learning and Deep Learning Techniques. Journal of Field Robotics, 40(1), 102-115. doi:10.1002/rob.22010

[15] Verma, S., & Singh, R. (2023). Deep Learning Approaches for Weed Detection in Soybean Fields. Expert Systems with Applications, 223, 120221. doi:10.1016/j.eswa.2023.120221

[16] Huang, Y., & Hu, Z. (2023). UAV-Based Weed Detection Using Deep Learning and Image Processing Techniques. Agricultural and Forest Meteorology, 314, 108717. doi:10.1016/j.agrformet.2023.108717

[17] Pappu, V., & Ganesan, B. (2023). Deep Learning for Weed Detection: A Case Study in Wheat Fields. Computers and Electronics in Agriculture, 198, 107036. doi:10.1016/j.compag.2022.107036

[18] Zhang, X., & Wu, Y. (2023). Interpretable Deep Learning Models for Weed Detection Using SHAP. IEEE Access, 11, 1023-1035. doi:10.1109/ACCESS.2023.1234567

[19] Das, S., & Mukherjee, A. (2023). Weed Detection Using Deep Convolutional Neural Networks in Rice Fields. Remote Sensing Applications: Society and Environment, 29, 100792. doi:10.1016/j.rsase.2023.100792

[20] Ramachandran, R., & Krishnan, M. (2023). Advances in Convolutional Neural Networks for Weed Detection in Agricultural Robotics. Robotics and Autonomous Systems, 163, 104393. doi:10.1016/j.robot.2023.104393.

[21] Dos Santos Ferreira, J. P., Freitas, D. P., da Silva, A. G., Pistori, H., & Folhes, M. T. (2017). Weed detection in soybean and carrot fields using convolutional neural networks. *Computers and Electronics in Agriculture*, 143, 314-324. https://doi.org/10.1016/j.compag.2017.11.027

[22] Olsen, A., Hanley, R., Zhang, C., McConchie, R., Knight, C., & Macdonald, B. (2019). DeepWeeds: A multiclass weed species image dataset for deep learning. *Frontiers in Plant Science*, 10, 1402. https://doi.org/10.3389/fpls.2019.01402

[23] Rußwurm, M., & Körner, M. (2018). Automated weed detection in maize fields with RGB images and deep learning. *Remote Sensing*, 10(2), 285. https://doi.org/10.3390/rs10020285

[24] Kang, L., Zhang, J., Liu, H., & He, Y. (2020). Real-time robust weed detection in the wild. *Computers and Electronics in Agriculture*, 170, 105252. https://doi.org/10.1016/j.compag.2020.105252

[25] Olsen, J., Hanly, R., Ball, D. A., & Hall, A. (2020). WeedNet: Dense semantic weed classification using multispectral images and MAV for smart farming. *Sensors*, 20(20), 5784. https://doi.org/10.3390/s20205784

[26] Bhosale, A., Bodkhe, S., Kumar, A., & Patel, S. (2021). Vision Transformers for dense prediction tasks: A study on weed segmentation. *IEEE Access*, 9, 126523-126534. https://doi.org/10.1109/ACCESS.2021.3111060

[27] Gupta, P., Singh, R., & Kumar, S. (2022). Weed detection in precision agriculture using Vision Transformers. *Computers and Electronics in Agriculture*, 193, 106591. https://doi.org/10.1016/j.compag.2022.106591