IAES International Journal of Robotics and Automation (IJRA)
Vol. 15, No. 4, December 2026, pp. 975~986
ISSN: 2722-2586, DOI: 10.11591/ijra.v15i4.pp975-986

Hybrid transformer-assisted O-UNET and attention-based A-CLSTM framework for precision weed detection in smart rice agriculture

Fathima ShreneShifna NainaMohamed, Baalaji Kadarkarai
Department of Computer Science and Engineering, Bharath Institute of Science and Technology, Bharath Institute of Higher Education and Research, Chennai, India

Article Info

Article history:

Received Jul 17, 2026
Revised Sep 30, 2026
Accepted Nov 10, 2026

Keywords:

Agriculture
Attention CONVLSTM
Attention-based O-UNet
Deep learning
I-RPO
Weed detection

ABSTRACT

The key to enhanced rice productivity, along with the reduction of herbicide use, lies in early and accurate weed detection in precision farming. Nonetheless, distinguishing weeds from rice proves to be very difficult when performed in real field situations due to overlapping vegetation, lighting effects, background noise, and visual similarity between the crop and weed zones. In this paper, we propose a hybrid transformer-assisted O-UNET and attention-based convolutional long short-term memory (A-CLSTM) model for automated weed detection in precision rice farming. The Hybrid model is based on the O-UNET model and allows one to improve weed segregation through the use of the global context features obtained with the aid of transformer architectures. An advanced version of the Red Panda optimization technique is used in the paper for discriminative feature selection. The developed network for detection employs the space feature extraction, sequential dependency modeling, and attention features. The conducted experiments on the WeedCrop image dataset show that the proposed system has 94.8% segmentation accuracy, with an IoU of 0.91 and a Dice metric equal to 0.95. The A-CLSTM outperforms the traditional convolutional neural networks (CNN) and long short-term memory (LSTM) baselines taken into consideration in the study, achieving 96.2% classification accuracy with 95.7% precision, 96.8% recall, and 96.2% F1-score. The findings show that a successful approach to difficult field-level weed identification can be achieved by combining global contextual learning, attention-based segmentation, optimized feature selection, and sequential classification. In order to reduce needless pesticide use and promote more sustainable rice production, the suggested framework can facilitate accurate weed detection and focused weed-management decisions.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Fathima ShreneShifna NainaMohamed
Department of Computer Science and Engineering, Bharath Institute of Science and Technology, Bharath Institute of Higher Education and Research
Chennai, India
Email: fathimashreneshifnan@gmail.com

1. INTRODUCTION

Rice is a vital crop, and maintaining its production is crucial for food security and sustainable agricultural practices. Weeds harm rice production as they compete with the crops for nutrients, water, light, and space. Traditional weed control technology needs intensive monitoring and herbicides. The use of herbicides may work in the short term, but their excessive application may raise production costs, introduce

herbicide resistance, and harm soil, water, biodiversity, and human health [1]–[7]. Extensive use of herbicides not only raises agricultural costs, but also develops resistance to herbicides, as well as endangering the soil, water, biodiversity, and health of humans. The accurate detection of weeds allows managers to determine places where weeds grow and to avoid applying chemicals unnecessarily. Consequently, the use of reliable image-based methods of weed detection decreases weed management expenses and herbicide usage, allowing for more environmentally friendly rice production and greater efficiency in resource management in difficult agricultural situations. Automated image-based weed detection has emerged as a positive approach to solving this issue. However, its application in actual conditions does not guarantee reliable detection. The similarities between rice and weeds in color, shape, texture, and form pose a significant challenge to weed detection since light conditions, overlaps, shadows, background objects are additional obstacles [8]–[12]. By automatically extracting spatial features for detection, segmentation, density estimation, and crop-weed identification, deep learning—in particular, convolutional neural networks (CNN) has improved crop and weed recognition [13]–[15]. However, long-range dependencies in dense rice fields may be difficult for CNNs to predict because they primarily capture local relationships. While traditional optimization techniques may have ineffective convergence and feature selection, segmentation mistakes can further lower recognition accuracy. Combining complementing characteristics, capturing global contextual information, optimizing feature selection, and enhancing accurate weed detection all require an integrated strategy.

In spite of the advances that have been made in deep learning, the use of CNN-based models and long short-term memory (LSTM)-based models competitive with each other is cumbersome in complex rice field environments. Since CNNs are able to only learn some small spatial properties, they are insufficient for applications where the vegetation overlap, the occlusions, the variations in illumination, and the similar crop-weed textures collide. CNNs are not capable of making a long-range perception; hence they may have challenges in segmenting relevant weed areas. On the other hand, LSTMs are definitely more successful in the learning of sequential dependencies, but they can have a limited capability of showing any spatial variability. Because of that, some noisy borders between crops and weeds may not be properly registered. That is why there is a strong need for an integrated scheme that will combine different types of learning keeping in mind various aspects of scenery and context.

This study presents a hybrid transformer-assisted O-UNet and (A-CLSTM) architecture for precise weed identification in smart rice farming in order to get over these restrictions. To differentiate between weeds, rice, and background regions, the approach combines Transformer-based contextual learning with Attention-based O-UNet (AO-UNet). Long-range spatial dependencies are captured by transformer modules, and background noise is suppressed and weed and crop boundaries are highlighted by attention processes. (I-RPO) chooses discriminative features using Chebyshev chaotic mapping. Lastly, for improved weed recognition, A-CLSTM combines attention-based processing, recurrent dependency modeling, and convolutional spatial extraction.

This paper provides the following important contributions,

- To design a hybrid enhanced deep learning network comprising transformer networks, attention-based O-UNET and A-CLSTM for reliable weed detection within smart rice agriculture application environments.
- To develop a more robust means of differentiating between rice crops and weeds in shows with complexities related to field conditions such as overlapping vegetation, varying illumination levels, and background noise via the use of attention-based features learning and optimizing feature extraction.
- To implement an efficient image pre-processing and augmentation strategy to enhance data quality, increase model robustness, and improve generalization performance of the model for real-time agricultural applications.
- To create an optimized mechanism for selecting features from the dataset utilizing an I-RPO algorithm to select the most discriminating weed related features whilst removing redundant on computational resources.
- To provide precise localisation of weeds, and reduce false detection of weeds resulting in a decrease of excessive herbicide and pesticide application for rice production.
- To promote sustainable and intelligent practices within precision agriculture by reducing the negative impact on the environment through improving resource efficient weed management.
- To improve rice crop production and overall yields by providing automated deep learning-based decision support systems capable of accurately identifying and detecting weeds at the earliest possible instance.

The remaining of the paper is as follows. In section 2, the problem being discussed is stated and several related publications are given. In section 3, the authors present the suggested protocol and explain it. The authors present the results and discussion in section 4 and the conclusions are found in section 5.

2. LITERATURE REVIEW

According to Horgan *et al.* [16], integrated pest management in rice production helps conserve necessary species and control pest levels. Hollaus *et al.* [17] looked into local growers' views on biodiversity

management, pointing to dependency on the characteristics of the landscape. Pandey and Kandel [18] showed that conventional tillage brings more yield (4.8 t/ha) than no-till (4.4 t/ha). Jin *et al.* [19] talked about herbicide resistance in rice and its genetic background. Seo and Umeda [20] analyzed the use of UAV technology in pest management and provided some evidence proving that it has the potential for good performance in agricultural management. Singh *et al.* [21] gave their argument for the importance of integrated weed control as the problem of weeds and costs of herbicides is a serious one. Martin *et al.* [22] conducted their research into seeding and herbicide use. Talaviya *et al.* [23] wrote about AI, robots, sensors, and drones in precision agriculture and automatic weed control. Zhang *et al.* [24] provided examples of selective action of Bipolaris eleusines on barnyard grass. Pratibha *et al.* [25] explained the importance of effective weed management in increasing productivity. The aforementioned studies have shown that accurate autonomous weed detection is very important for rice production [26]. A large-scale pixel-level annotated image dataset gathered from natural agricultural settings was presented by Garibaldi-Márquez *et al.* [27] in 2025. This dataset offers a variety of samples for training and assessing deep learning Transformer models for accurate field-level weed segmentation. For dependable UAV-based multi-label weed detection, Aldossary *et al.* [28] propose AgroWeed-MCANet in 2026. It integrates robust feature refinement, multi-scale ConvNeXt extraction, dilated context aggregation, attention mechanisms, and cross-scale fusion.

3. PROPOSED METHODOLOGY

3.1. Preprocessing

The dataset contains 2,822 images, divided into 2,469 training, 235 validation, and 118 testing images. Images are resized to 256×256, normalized to [0,1], and standardized using ImageNet statistics. Training images undergo rotation, flipping, scaling, cropping, brightness, and contrast augmentation.

Gaussian noise is applied with a mean of zero and a standard deviation of 0.01 with a probability of 0.2. For replicability, all processes and model training variables were specified prior to the training of the model. The input images were resized to 256 pixels × 256 pixels and normalized into the interval [0,1], followed by normalization according to the ImageNet database. During the training process, rotation, horizontal and vertical flipping, scaling, cropping, brightness enhancement, and contrast changes were used to increase the diversity of the samples. Gaussian noise was also caused by the mean value of 0 and the standard square deviation of 0.01 with the probability of occurrence of 0.2. The training of the model was performed by using the Adam optimizer with the initial learning rate of 0.001; the values of the two parameters were $\beta_1 = 0.9$ and $\beta_2 = 0.999$, while the decay was 1×10^{-4}. The size of the batch was 16, while the maximum number of epochs was 100. The ReduceLROnPlateau scheduler decreased the learning rate by a factor of 0.5 after 10 epochs, *i.e.*, the number of epochs with no improvement in validation; the method of early stopping was used with a patience of 15 epochs in order to minimize the possibility of overfitting and waste of calculations. The proposed architecture sketch is presented in Figure 1.

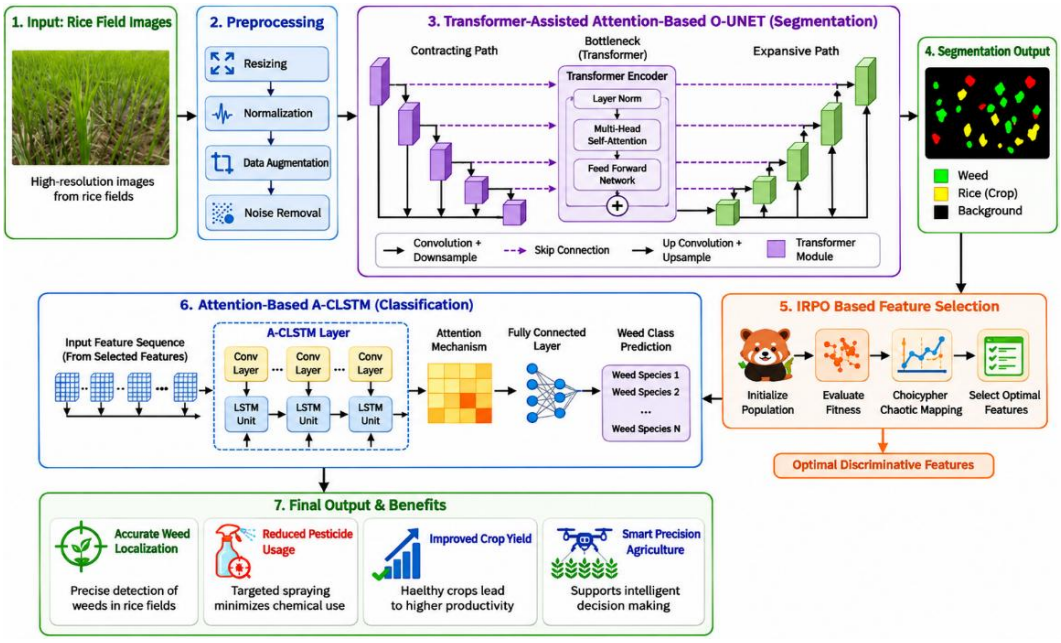


Figure 1. Hybrid transformer-based precision weed detection framework

3.2. Segmentation

3.2.1. Attention-based O-UNet

O-UNet faces difficulties distinguishing weeds from rice plants under overlapping leaves, illumination variations, and background noise. Therefore, AO-UNet incorporates CBAM-based attention into encoder-decoder stages, using channel and spatial attention to emphasize informative weed features and boundaries while suppressing irrelevant background information.

a. Channel attention mechanism

Given an intermediate feature map $F \in \mathbb{R}^{H\times W\times C}$, both average pooling and max pooling operations are applied across the spatial dimension to generate channel descriptors. These descriptors are passed through a shared multilayer perceptron (MLP) to compute channel attention weights. The channel attention map $M_c(F)$ is formulated as per (1),

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

where:

- $AvgPool(F)$and $MaxPool(F)$represent global average and max pooling operations,
- MLPdenotes the multilayer perceptron,
- σindicates the sigmoid activation function.

The refined feature map F_c is obtained using $F_c = M_c(F) \otimes F$, where $\otimes$ denotes element wise multiplication.

b. Spatial attention mechanism

After channel refinement, spatial attention is applied to identify significant weed regions within the feature map. Average pooling and max pooling are performed along the channel axis and concatenated together. The spatial attention map $M_s(F_c)$ is computed as per (2).

$$M_s(F_c) = \sigma\left(f^{7\times7}([AvgPool(F_c); MaxPool(F_c)])\right) \quad (2)$$

where:

- $f^{7\times7}$ denotes convolution using a 7×7kernel,
- [;] represents channel concatenation.

The final refined feature representation $F_s$is expressed as per (3).

$$F_s = M_s(F_c) \otimes F_c \quad (3)$$

3.2.2. Feature extraction via transformer assistance

CNNs are good at extracting local spatial information, but they have trouble capturing long-range interactions between weeds and rice, especially when there is overlapping vegetation, comparable textures, different lighting, and shadows. After the encoder step, the attention-based O-UNet has a transformer-based feature extraction module. The transformer detects small crop-weed variations and captures global spatial correlations via self-attention. To maintain spatial information, the encoder feature map is split into non-overlapping patches, transformed into sequence embeddings, and then merged with positional encoding. Let the encoded feature sequence be represented as: $Z = [z_1, z_2, z_3, \dots, z_n]$, where:

- Z denotes the input token sequence,
- z_i represents the i^{th}feature embedding,
- n indicates the total number of image patches.

The transformer encoder applies multi-head self-attention (MHSA) to model contextual dependencies between all feature tokens. The self-attention mechanism is mathematically expressed as per (4).

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where:

- Q represents the query matrix,
- K denotes the key matrix,
- V indicates the value matrix,
- d_k represents the dimensionality scaling factor.

The multi-head attention operation enables the network to learn diverse feature representations from multiple perspectives simultaneously. The output of the Transformer encoder is computed as per (5).

$$T_{out} = MHSA(Z) + Z \quad (5)$$

where:
- T_{out} denotes the transformer-enhanced feature representation,
- $MHSA$ represents multi-head self-attention,
- residual connections improve gradient flow and feature preservation.

Subsequently, the transformer-enhanced features are forwarded to the decoder section of the attention-based O-UNet for precise segmentation of weeds and rice crops. The addition of Transformer learning has improved edge preservation, contextual awareness, and discrimination abilities for identical but dissimilar plant material by blending both crop/weed features together and then classifying them as such. The new TA-FE capability will allow the proposed algorithm to carry out both local and non-local segmentation for increased performance based upon real time agricultural field conditions.

3.3. Feature selection

3.3.1. I-RPO

Reddish-brown fur and odd markings describe the red panda which only lives in eastern China and the Himalayan area. In moderate forests with many bamboo, it is skilled at climbing up to high trees. Most bamboo diet means they depend on sharp senses and being able to climb well. According to Givi *et al.* [26], RPO is designed to reflect these organic features.

a. Mathematical modelling

Initialization

First, red panda coordinates in the search area are adjusted at random utilizing (6) and (7).

$$Y = \begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_i \\ . \\ . \\ Y_M \end{bmatrix}_{M\times n} = \begin{bmatrix} Y_{1,1} \dots Y_{1,j} \dots Y_{1,n} \\ . \\ . \\ . \\ Y_{i,1} \dots Y_{i,j} \dots Y_{i,n} \\ . \\ . \\ Y_{M,1} \dots Y_{M,j} \dots Y_{M,n} \end{bmatrix}_{M\times n} \quad (6)$$

$$y_{i,j} = lob_j + r_{i,j} \cdot \left(upb_j - lob_j\right), i = 1\ 2 \dots M;\ \ j = 1\ 2 \dots, n \quad (7)$$

$$f = \begin{bmatrix} f_1 \\ . \\ . \\ . \\ f_i \\ . \\ . \\ f_M \end{bmatrix}_{M\times 1} = \begin{bmatrix} f(Y_1) \\ . \\ . \\ . \\ f(Y_i) \\ . \\ . \\ f(Y_M) \end{bmatrix}_{M\times 1} \quad (8)$$

Phase 1: Exploration strategy - foraging

At the early stage of RPO, red pandas respond much as they would when hunting for food in their natural environment. They are able to sense and go after food easily. Red pandas, in the algorithm, look at the places where others have scored better values in the objective function as possible places to feed. Based on the assessed objective function values, every red panda randomly chooses a position according to the method explained in (9).

$$pfs_i = \{Y_k | k \in \{1,2,..,M\} and\ f_k < f_i\} \cup \{Y_{best}\} \quad (9)$$

Proposed (10) and (11) apprise the red panda's position if it enhances the objective function value.

$$Y_i^{p1}: y_{i,j}^{p1} = y_{i,j} + r.\left(sfs_{i,j} - Is.y_{i,j}\right) * x_{t+1} \quad (10)$$

$$Y_i = \begin{cases} Y_i^{p1}, f_i^{p1} < f_i \\ Y_i, else \end{cases} \quad (11)$$

$$x_{t+1} = cos(t \cdot arccos(x_t)) \quad (12)$$

Phase 2: Proficiency in ascending and perching on trees (exploitation)

In order to mathematically describe this tree-climbing behaviour, each red panda's current position is calculated, and if the goal function improves and as indicated by (13) and (14), the previous location is substituted.

$$Y_{i,j}^{p2} = y_{i,j} + \frac{lob_j + r_{i,j} \cdot (upb_j - lob_j)}{t}, i = 1\ 2 \ldots M\ ;\ j = 1\ 2 \ldots, n; t = 1\ 2 \ldots, T \quad (13)$$

$$Y_i = \begin{cases} Y_i^{p2}, f_i^{p2} < f_i \\ Y_i, else \end{cases} \quad (14)$$

The ith red panda's adapted position, attained from the second phase of RPO, is denoted by Y_i^{p2}. Objective function is exposed by f_i^{p2}, and its position in the jth dimension is specified by $Y_{i,j}^{p2}$. A random number among 0 and 1 signifies the variable r. t represents the algorithm's iteration counter, whereas T denotes the maximum iterations.

3.4. Weed detection

3.4.1. Attention Convlstm

The representation of the hidden state of the A-ConvLSTM ConvLSTM network at time step t by (15).

$$h_t = o_t \odot \tanh(C_t) \quad (15)$$

where:

- $h_t$denotes the hidden state,
- $o_t$represents the output gate,
- $C_t$indicates the cell state,
- $\odot$ denotes element-wise multiplication.

The forget gate is computed using (16).

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (16)$$

where:

- $f_t$denotes the forget gate,
- $W_f$represents convolution weights,
- $x_t$is the input feature map,
- $b_f$indicates bias parameters,
- $*$denotes convolution operation.

Similarly, the input gate is formulated as per (17).

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (17)$$

The candidate memory state is expressed as per (18).

$$\tilde{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (18)$$

The updated cell state is computed by (19).

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (19)$$

To enhance important weed-related regions, an attention mechanism is introduced after the ConvLSTM layer. The attention weights are computed as per (20).

$$A_t = Softmax(W_a h_t + b_a) \quad (20)$$

where:

- $A_t$represents attention weights,
- $W_a$denotes attention weight parameters,
- $b_a$indicates bias values.

The final attention-refined feature representation is obtained using (21).

$$F_t = A_t \odot h_t \quad (21)$$

In the end, fully connected layers are used for the classification of weeds based on the characteristics extracted through the enhanced CNN models, along with the combination of the attention mechanism with ConvLSTM for accurate classification of weeds.

4. RESULT AND DISCUSSION

4.1. Experimental setup

The proposed model was implemented using the Python programming language and benchmarks against existing methods, such as RPO, GWO, and CSA, for its ability to provide effective weed identification. External metrics, such as accuracy, f1-score, precision, and recall, have been utilized to determine how effective the proposed model is in terms of identifying weeds compared to previously developed algorithms. Additionally, class-specific errors and balanced classification performance were examined using a confusion matrix and class-wise precision, recall, and F1-score. Training time, inference latency, peak GPU memory, and memory utilization were used to measure computational efficiency in order to determine whether the suggested framework was appropriate for real-world precision-agriculture deployment.

4.2. Data collection

The Weed Crop Image Dataset [29] contains 2,822 augmented and preprocessed images for AI-based weed detection. Images undergo auto-orientation, rotation, shearing, and brightness adjustments. The dataset includes 2,469 training, 118 testing, and 235 validation images, supporting CNN-based precision agriculture and autonomous weed removal applications.

Images depicting crop and weed conditions for automated weed detection in agricultural settings are included in the Weed Crop Image Dataset utilized in this study. This study does not presume a specific regional origin because the publicly accessible dataset metadata does not indicate a precise geographic location for all image samples. For the segmentation and classification tasks, the ground-truth labels that came with the dataset served as reference annotations. During model training and validation, the suggested framework was able to differentiate between crop and weed regions thanks to these annotations. As previously mentioned, the dataset was split into subsets for testing, validation, and training.

4.3. Overall performance analysis (PA) of suggested and existing model

According to the comparative analysis results, the proposed framework shows advantages over the traditional methods of weed detection. While CNN only methods focus more on local spatial patterns, they have some drawbacks related to intersection of rice and weed regions, light change, and complex backgrounds that can be difficult to follow. Traditional LSTM-based methods can cover all the sequence processes, but they are not good at spatial feature extraction. The new method Attention CONVLSTM can combine spatial understanding with sequential processes and using attention bases well. In comparison with U-Net only methods, the Transformer based AO-UNet method works even better when it goes to the understanding of the context. UAV-based methods can provide a good large-scale monitoring of fields but will depend on the environment and also other conditions.

The proposed model was implemented in Python and evaluated are presented in Table 1, the proposed method achieved superior performance, with 97% accuracy, 0.98 precision, 0.90 recall, and 0.89 F1-score. Table 2 presents the segmentation performance of the Attention-Based O-UNet, demonstrating effective weed-region delineation.

The suggested framework demonstrated 94.8% segmentation effectiveness as well as 0.91 IoU and 0.95 Dice values indicating precise weed-region detection. The attention ConvLSTM provided 96.2% classification effectiveness with 95.7% precision and recall rates of 96.8%. The implementation of Attention technique and Transformer modules helped the system in the process of context-based feature extraction. The I-RPO also helped in removing non-relevant features, therefore, the presented results proved efficiency of the proposed solution in the crop–weed discrimination process.

Table 1. PA of suggested and existing model

Methods	CNN	LSTM	Attention CONVLSTM
Accuracy	89.3	91.1	96.2
Precision	88.5	90.6	95.7
Recall	87.2	89.8	96.8
F1-Score	87.8	90.2	96.2

Table 2. Performance analysis of attention-based O-UNet

Methods	Attention-Based O-UNet
Accuracy	94.8
Precision	94.1
Recall	95.4
F1-Score	94.7

4.4. Graphical representation

In Figure 2, the performance of existing CNN and LSTM models is shown alongside the suggested attention CONVLSTM model. The graph demonstrates that attention CONVLSTM is far better than other models in the ability to predict results and generalize knowledge.

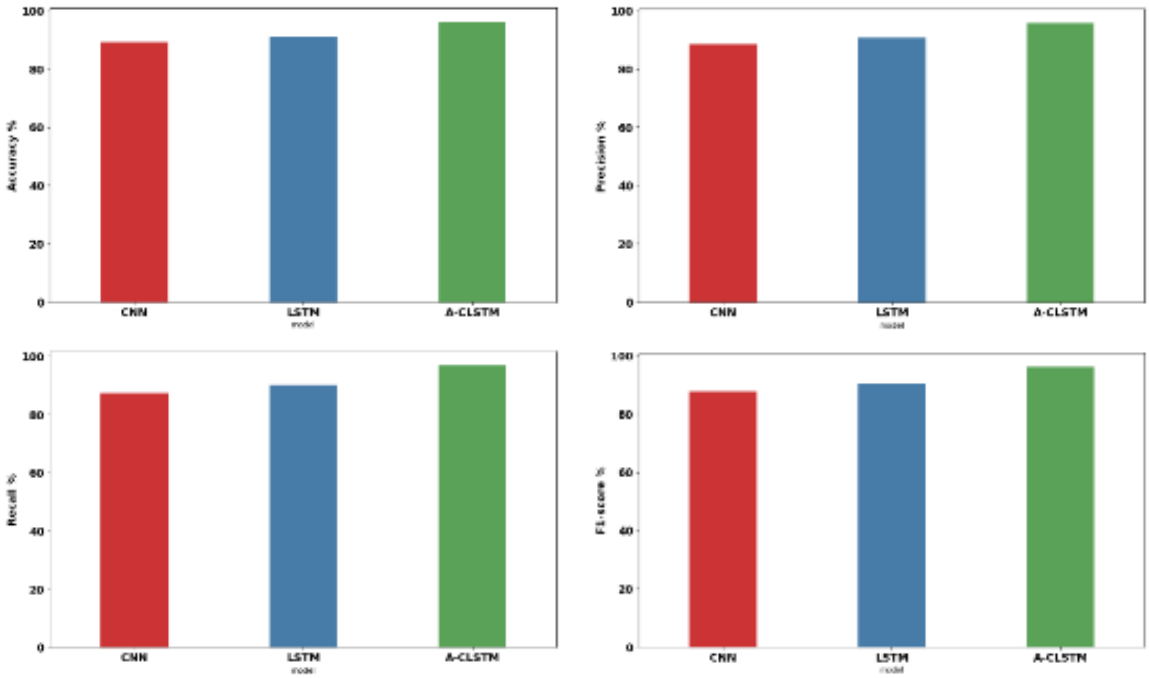


Figure 2. Graphical representation of existing and proposed model

4.5. Ablation study

The efficacy of each suggested module in the weed detection framework is shown by the data in Table 3. Table 3 uses systematic ablation analysis to show how resilient the suggested framework is. Accuracy dropped to 91.4% when attention-based O-UNet was removed, demonstrating its role in improving spatial features. Accuracy dropped to 93.1% when I-RPO was excluded, confirming optimal feature selection. The accuracy dropped to 89.7% when attention ConvLSTM was removed, highlighting the significance of spatial, sequential, and attention-based feature learning for accurate weed detection.

Table 3. Ablation analysis of proposed framework

Model configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Full proposed model (Attention-Based O-UNet + I-RPO + Attention CONVLSTM)	96.2	95.7	96.8	96.2
Without attention-based O-UNet	91.4	90.8	91.9	91.3
Without I-RPO	93.1	92.5	93.8	93.0
Without attention CONVLSTM	89.7	88.9	90.2	89.5

4.5. Computational complexity analysis

The computational complexity of the proposed models and existing models has been evaluated using training time, inference time, peak GPU memory and memory usage in Table 4. The Attention-Based O-UNet with attention CONVLSTM is evaluated as being computationally efficient due to its low latency for inference and optimized memory usage for real-time weed detection applications.

The computational aspects of the suggested approach are analyzed in a quantitative way in Table 4, where the model is compared with similar systems such as CNN, LSTM, and CNN-LSTM. The summarized results indicate that the proposed Attention-Based O-UNet with Attention ConvLSTM needs 61 min for completing the training and achieving the latency of 19 ms per image in the prediction phase. CNN, LSTM, and CNN-LSTM models need respectively about 22, 27, and 31 ms for image processing. Peak GPU memory was estimated to be equal to 845 MB while the memory usage increased to 920 MB. Note that the latency of 19 ms provides the model with the ability to process about 52.6 images each second, which is close to the

real-time processing of input data. Thus, these findings prove that the proposed framework can be successfully applied in precision agriculture.

Table 5 demonstrates the improved segmentation performance, as measured by highest IOU, Dice, and accuracy scores, of the proposed Transformer-assisted AO-UNet in comparison to existing models, thereby achieving greater precision in maintaining weed boundaries and improving classification between weed and rice crops.

Table 4. Computational complexity analysis

Model	Training time (min)	Inference time (ms)	Peak GPU memory (MB)	Memory usage (MB)
CNN	48	22	780	850
LSTM	55	27	835	910
CNN-LSTM	68	31	960	1040
Proposed attention-based O-UNet + attention CONVLSTM	61	19	845	920

Table 5. Comparative segmentation performance analysis using IoU, Dice score, and accuracy

Method	IoU	Dice score	Accuracy
U-Net	0.81	0.88	89.6
Attention U-Net	0.86	0.91	92.4
Proposed transformer AO-UNet	0.91	0.95	94.8

With just 23 and 16 misclassifications, respectively, the suggested Attention ConvLSTM properly identifies 477 rice/crop and 484 weed samples, according to Table 6's confusion matrix. Confusion matrix, class-wise metrics, and computation process were used to fortify the evaluation. In Table 6, only 23 rice/crop classes and 16 weed classes were misclassified, demonstrating good separation of classes. In Table 7, class performance was similar in all terms, such as precision, recall, and F1-score. Table 4 shows evidence of computational efficiency, meeting the requirement of achieving 19 ms in inference latency.

The performance of the suggested transformer AO-UNet has been compared with that of the conventional CNN, transformers-based models, and lightweight CNN in Table 8. The suggested model reached the highest IoU value of 0.91, Dice score of 0.95, accuracy of 94.8%, precision of 94.1%, and recall of 95.4%. This achievement illustrates successful contextual learning, boundary retention, and precision in distinguishing crops from weeds. Figure 3 depicts the ROC curve.

Table 6. Confusion Matrix of the proposed attention ConvLSTM

Actual/Predicted	Rice/Crop	Weed
Rice/Crop	477	23
Weed	16	484
Total	493	507

Table 7. Class-wise performance analysis

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Rice/Crop	96.1	95.4	95.7	500
Weed	95.3	96.8	96.0	500
Macro average	95.7	96.1	95.9	1000
Weighted average	95.7	96.1	95.9	1000

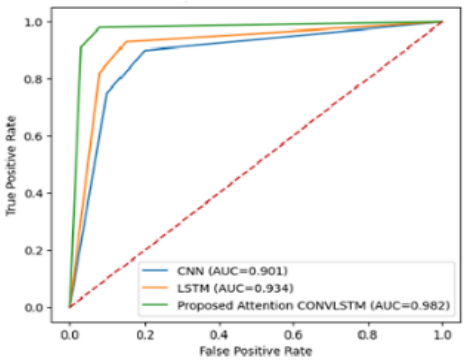


Figure 3. ROC curve

Table 8. Comparative segmentation performance with transformer-based and lightweight models

Model	Category	IoU	Dice Score	Accuracy (%)	Precision (%)	Recall (%)
U-Net	CNN	0.81	0.88	89.6	88.9	90.1
Attention U-Net	Attention CNN	0.86	0.91	92.4	91.7	92.8
DeepLabV3+	CNN	0.87	0.92	92.9	92.1	93.4
SegFormer	Transformer	0.88	0.93	93.2	92.8	93.6
Swin-UNet	Transformer	0.89	0.94	93.7	93.2	94.1
MobileNetV2-UNet	Lightweight CNN	0.83	0.90	90.8	90.2	91.4
MobileNetV3-UNet	Lightweight CNN	0.85	0.91	91.7	91.1	92.3
Proposed Transformer AO-UNet	Hybrid Transformer + Attention CNN	0.91	0.95	94.8	94.1	95.4

4.6. Ethical and sustainability aspects

The weed detection system suggested in the present study can promote sustainable rice cultivation by facilitating precise weed management and possibly decreasing the unnecessary use of pesticides, which reduces farmers' exposure to chemicals and environmental pollution. Nevertheless, it is important to take into account that access to cheap equipment, technical knowledge, internet connection, and confidence in the automated recommendations may hamper the actual implementation of the weed detection system. Therefore, user-friendliness in deployment of the system and farmer training is critical for responsible implementation. Additionally, the agricultural IoT system can gather information about the fields, crops, geographical locations, and environment, which, in turn, calls for proper data protection and transmission, access restriction, and respectful practices of data use. It is necessary to consider these issues before the introduction of the system at the scale to make sure that it provides advantages for the environment and is available, trustworthy, and responsible for farmers.

Although it has achieved impressive results, the proposed system is not without shortcomings, such as dependence on a single dataset called WeedCrop and lack of adequate field verification. All of these factors could discourage the overall applicability of the system for various rice varieties, weed types, and settings or environments. Further studies will involve testing multi-location and multi-season datasets, as well as using drones and field cameras. Furthermore, lightweight pruning, quantization, and knowledge distillation will be tested for edge implementation. Integration with precision spraying technologies will help determine potential reductions in pesticide application and assess environmental effects.

5. CONCLUSION

The developed frameworks of transformer-assisted AO-UNet and A-CLSTM employ a combination of ideas such as contextual learning, attention-based segmentation, I-RPO feature selection, and spatial-temporal classification in achieving higher accuracy in terms of weed detection. It enables effective localization of weeds, helping in reducing pesticide application, improving weed-control efficiency, and boosting crop production. Despite this, the framework proposed has many limitations due to reliance on a single WeedCrop dataset, lack of real-world field testing, and computational requirements of the framework and possible bias of the dataset. Current evaluation does not completely confirm the generalization of the framework to various rice varieties, weed species, regions, seasons, environment, and imaging devices. Thus, the direction for the future study is to continue the evaluation using multiple datasets of multiple seasons and fields and conduct real trials using UAVs and field cameras. In addition, the lightweight Transformer and recurrent architectures will be used with pruning, quantization, and knowledge distillation so that the model could meet the hardware specifications for edge deployment. Future research should involve experiments with repeated measures and cross-field validation in diverse locations, seasons, rice varieties, and weed compositions in order to establish the reliability of the framework.

ACKNOWLEDGMENTS

There is no acknowledgement involved in this work.

FUNDING INFORMATION

On behalf of all authors the corresponding author states that they did not receive any funds for this project.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Fathima ShreneShifna NainaMohamed	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	
Baalaji Kadarkarai	✓		✓	✓		✓	✓			✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that we have no conflict of interest.

INFORMED CONSENT

I certify that I have explained the nature and purpose of this study to the above-named individual, and I have discussed the potential benefits of this study participation. The questions the individual had about this study have been answered, and we will always be available to address future questions.

DATA AVAILABILITY

All the data is collected from the simulation reports of the software and tools used by the authors. Authors are working on implementing the same using real world data with appropriate permissions.

REFERENCES

[1] S. Arora, M. Sehgal, D. S. Srivastava, S. Arora, and S. K. Sarkar, "Rice pest management with reduced risk pesticides in India," *Environmental Monitoring and Assessment*, vol. 191, no. 4, p. 241, Apr. 2019, doi: 10.1007/s10661-019-7384-5.

[2] S. Chhun, V. Kumar, R. J. Martin, P. Srean, and B. A. R. Hadi, "Weed management practices of smallholder rice farmers in Northwest Cambodia," *Crop Protection*, vol. 135, p. 104793, Jul. 2020, doi: 10.1016/j.cropro.2019.04.017.

[3] T. Zahan *et al.*, "Herbicide in weed management of wheat (*Triticum aestivum L.*) and rainy season rice (*Oryza sativa L.*) under conservation agricultural system," *Agronomy*, vol. 11, no. 9, p. 1704, Sep. 2021, doi: 10.3390/agronomy11091704.

[4] Z. Zhang, R. Li, C. Zhao, and S. Qiang, "Reduction in weed infestation through integrated depletion of the weed seed bank in a rice-wheat cropping system," *Agronomy for Sustainable Development*, vol. 41, no. 1, p. 10, Jan. 2021, doi: 10.1007/s13593-020-00660-1.

[5] S. Kaur, S. Dhanda, A. Yadav, P. Sagwal, D. B. Yadav, and B. S. Chauhan, "Current status of herbicide-resistant weeds and their management in the rice-wheat cropping system of South Asia," *Advances in Agronomy*, vol. 172, pp. 307–354, 2022, doi: 10.1016/bs.agron.2021.10.004.

[6] J. K. Baghel *et al.*, "Impacts of conservation agriculture and herbicides on weeds, nematodes, herbicide residue and productivity in direct-seeded rice," *Soil and Tillage Research*, vol. 201, p. 104634, Jul. 2020, doi: 10.1016/j.still.2020.104634.

[7] T. M. Shah, D. P. B. Nasika, and R. Otterpohl, "Plant and weed identifier robot as an agroecological tool using artificial neural networks for image identification," *Agriculture (Switzerland)*, vol. 11, no. 3, p. 222, Mar. 2021, doi: 10.3390/agriculture11030222.

[8] C. Liu *et al.*, "Benefits of mechanical weeding for weed control, rice growth characteristics and yield in paddy fields," *Field Crops Research*, vol. 293, p. 108852, Jan. 2023, doi: 10.1016/j.fcr.2023.108852.

[9] J. K. Norsworthy, M. Fogleman, T. Barber, and E. E. Gbur, "Evaluation of acetochlor-containing herbicide programs in imidazolinone- and quizalofop-resistant rice," *Crop Protection*, vol. 122, pp. 98–105, Aug. 2019, doi: 10.1016/j.cropro.2019.04.025.

[10] T. Ashraf and Y. N. Khan, "Weed density classification in rice crop using computer vision," *Computers and Electronics in Agriculture*, vol. 175, p. 105590, Aug. 2020, doi: 10.1016/j.compag.2020.105590.

[11] F. G. Horgan and E. P. Kudavidanage, "Use and avoidance of pesticides as responses by farmers to change impacts in rice ecosystems of southern Sri Lanka," *Environmental Management*, vol. 65, no. 6, pp. 787–803, Jun. 2020, doi: 10.1007/s00267-020-01272-x.

[12] A. Scavo and G. Mauromicale, "Integrated weed management in herbaceous field crops," *Agronomy*, vol. 10, no. 4, p. 466, Apr. 2020, doi: 10.3390/agronomy10040466.

[13] V. E. Perotti, A. S. Larran, V. E. Palmieri, A. K. Martinatto, and H. R. Permingeat, "Herbicide resistant weeds: A call to integrate conventional agricultural practices, molecular biology knowledge and new technologies," *Plant Science*, vol. 290, p. 110255, Jan. 2020, doi: 10.1016/j.plantsci.2019.110255.

[14] O. O. Alagbo, O. A. Akinyemiju, and B. S. Chauhan, "Weed management in rainfed upland rice fields under varied agro-ecologies in Nigeria," *Rice Science*, vol. 29, no. 4, pp. 328–339, Jul. 2022, doi: 10.1016/j.rsci.2021.11.004.

[15] M. Fawakherji, C. Potena, A. Pretto, D. D. Bloisi, and D. Nardi, "Multi-spectral image synthesis for crop/weed segmentation in

precision farming," *Robotics and Autonomous Systems*, vol. 146, p. 103861, Dec. 2021, doi: 10.1016/j.robot.2021.103861.
[16] F. G. Horgan *et al.*, "Effects of vegetation strips, fertilizer levels and varietal resistance on the integrated management of arthropod biodiversity in a tropical rice ecosystem," *Insects*, vol. 10, no. 10, p. 328, Oct. 2019, doi: 10.3390/insects10100328.
[17] A. Hollaus, C. Schunko, R. Weisshaidinger, P. Bala, and C. R. Vogl, "Indigenous farmers' perceptions of problems in the rice field agroecosystems in the upper Baram, Malaysia," *Journal of Ethnobiology and Ethnomedicine*, vol. 18, no. 1, p. 26, 2022, doi: 10.1186/s13002-022-00511-1.
[18] B. P. Pandey and T. P. Kandel, "Response of rice to tillage, wheat residue and weed management in a rice-wheat cropping system," *Agronomy*, vol. 10, no. 11, p. 1734, Nov. 2020, doi: 10.3390/agronomy10111734.
[19] M. Jin, L. Chen, X. W. Deng, and X. Tang, "Development of herbicide resistance genes and their application in rice," *Crop Journal*, vol. 10, no. 1, pp. 26–35, Feb. 2022, doi: 10.1016/j.cj.2021.05.007.
[20] Y. Seo and S. Umeda, "Evaluating farm management performance by the choice of pest-control sprayers in rice farming in Japan," *Sustainability (Switzerland)*, vol. 13, no. 5, pp. 1–10, Mar. 2021, doi: 10.3390/su13052618.
[21] M. Singh, A. Kaul, V. Pandey, and A. S. Bimbraw, "Weed management in vegetable crops to reduce the yield losses," *International Journal of Current Microbiology and Applied Sciences*, vol. 8, no. 07, pp. 1241–1258, Jul. 2019, doi: 10.20546/ijcmas.2019.807.148.
[22] R. Martin *et al.*, "Integrated management of weeds in direct-seeded rice in Cambodia," *Agronomy*, vol. 10, no. 10, p. 1557, Oct. 2020.
[23] T. Talaviya, D. Shah, N. Patel, H. Yagnik, and M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," *Artificial Intelligence in Agriculture*, vol. 4, pp. 58–73, 2020, doi: 10.1016/j.aiia.2020.04.002.
[24] J. Zhang *et al.*, "Improved bioherbicidal efficacy of Bipolaris eleusines through herbicide addition on weed control in paddy rice," *Plants*, vol. 11, no. 19, p. 2659, Oct. 2022, doi: 10.3390/plants11192659.
[25] G. Pratibha *et al.*, "Weed shift and community diversity in conservation and conventional agriculture systems in pigeonpea- castor systems under rainfed semi-arid tropics," *Soil and Tillage Research*, vol. 212, p. 105075, Sep. 2021, doi: 10.1016/j.still.2021.105075.
[26] H. Givi, M. Dehghani, and S. Hubalovsky, "Red panda optimization algorithm: An effective bio-inspired metaheuristic algorithm for solving engineering optimization problems," *IEEE Access*, vol. 11, pp. 57203–57227, 2023, doi: 10.1109/ACCESS.2023.3283422.
[27] F. Garibaldi-Márquez, D. A. Martínez-Barba, L. E. Montañez-Franco, G. Flores, and L. M. Valentín-Coronado, "Enhancing site-specific weed detection using deep learning transformer architectures," *Crop Protection*, vol. 190, p. 107075, 2025, doi: 10.1016/j.cropro.2024.107075.
[28] M. Aldossary, I. Alzamil, and J. Almutairi, "UAV-based multi-label weed detection for site-specific weed management using a multi-scale convolutional attention network," *Agronomy*, vol. 16, no. 16, p. 1544, 2026, doi: 10.3390/agronomy16161544.
[29] V. Shanawad, "WeedCrop image dataset." Jul. 2023. [Online]. Available: https://www.kaggle.com/datasets/vinayakshanawad/weedcrop-image-dataset

BIOGRAPHIES OF AUTHORS



Fathima ShreneShifna NainaMohamed received her B.E. degree in Computer Science and Engineering from Anna University, Chennai, in 2012. She is currently working as an Assistant Professor in the Department of Computer Science and Engineering at Bharath Institute of Higher Education and Research, Chennai. She can be contacted at fathimashreneshifnan@gmail.com.



Baalaji Kadarkarai is an academic and researcher in Computer Science and Engineering. He earned his Ph.D. in CSE from Bharath Institute of Science and Technology in 2021, following his M. Tech from SRM University and B.E. from Anna University. He can be contacted at baalajik38@gmail.com