An unique model for weed and paddy detection using regional convolutional neural networks

M. Vaidhehi and C. Malathy
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, India

ABSTRACT

Aim: In the agricultural field, weeds are grown irrespective of the required species, which spoils the growth of paddy plants. The presence of weeds is to be detected and should be classified in the earlier stage to improve the growth of species. This research work considers paddy cultivation and detection of weeds in the paddy field.

Methods: The modelling of the automatic weed predictor model aids farmers in handling the weed coverage and scattering of weed in the agricultural field. Real-time data is collected from the agricultural region, and the images are provided as the input for the predictor model. Regional Convolutional Neural Networks (R-CNN) is proposed to segment the weed from the input images.

Results: The model is proposed to address the segmentation problem by concurrent simulation of the task for object prediction. Simulation is carried out in a MATLAB environment. The performance of R-CNN is compared and evaluated with existing approaches like the conventional CNN model and other segmentation approaches.

Conclusion: The proposed model gives better results when compared to other approaches.

Introduction

The cultivation of paddy is considered as an essential part of agriculture all over the world. The pests and weeds are the foremost cause of harm which reduces the crop quality towards the agricultural process (Thenmozhi and Reddy 2019). The growths of crops are affected due to the parallel growth of weeds at the various crop growth stages. It grabs the nutrition from the sun and water and competes with the crop, and destroys the quality of cultivation. The adoption of weed control methods helps in reducing the loss to a major percentage and reduces the occurrence of other crop-based diseases. Generally, the farmers adopt weed control methods, i.e. non-chemical and chemical techniques used extensively over the cultivation field (Cheng et al. 2017). The farmer model is time-consuming, costly, and a tedious process. The substitution to this method is the mechanical control methods. Various conventional approaches are utilised to manage the growth of pests and weeds for attaining better yields (Chauhan et al. 2012). The drawbacks of these approaches are crop contamination and environmental pollution, which have a major impact on human health. The advancements in machine vision technologies offer a professional tool for weed tracing using a mechanical weeding system.

The researchers good crop discrimination and plant detection from weeds are considered as a major challenge encountered by the researchers. Chauhan and Opena (2013) anticipate a novel approach for feature extraction by examining the morphological characteristics of the paddy from the crop images. Robots are also equipped with screw-type wheels for weeding purposes. Tang et al. (2017) modelled a novel robotic design for controlling weed over the rice field. The model possesses a stereo camera and laser range finder for the robotic arm and autonomous mobility for controlling the weed. The obstacles in the prediction of rice plants and weeds during crop growing stages and any harm towards the cultivation is a most confronting task. However, the execution of robotic weeding towards the conventional row-based cropping field and planting schemes requires hardware technologies and complex software.

Some investigators adopt a chemical methodology that specifies the traditional way to eradicate weed and offer superior weed control mechanisms based on herbicides applications. Various investigators also analyse the significance of the chemical weed control
methods towards crop yield. The adoptions of chemical weed control mechanism for paddy cultivation encounters various disadvantages. It is significant due to the conventional spraying process, which causes added herbicides and leads to crucial problems like safety and ecological measures, environmental pollution, and so on (Gaire et al. 2013). It projects the significance of exploring sustainable methods for reducing the utilisation of agro-chemicals by retaining the efficacy. Some weed detection approaches with precision spraying are more appropriate for optimising herbicide usage and reducing the factors that impact the environment.

Presently, the wider advancements in automatic expert systems and computer vision have provided fascinating outcomes to different weeds and crops in an easier manner. Some ground-based weed detection process is adopted for certain site-specific and herbicide applications for weed management as eco-friendly approaches (Joshi et al. 2015). It helps to reduce the chemical pesticide consumption and environmental consequences over the farms. For this cause, various researchers adopt remote sensing approaches to discriminate soil, crops and weeds over reflective measurements at various wavelengths. Simultaneously, other investigators concentrate on modelling image processing approaches for classifying plants from the weeds based on texture, morphological, and colour features (Khaliq et al. 2014). Machine learning-based image processing is applied to two diverse forms of images with 2D (two-dimensional) and 3D (three-dimensional). It is based on 2D image processing possesses drawbacks with 2D cameras. Initially, variations in the outdoor environment (illumination) are captured by the cameras; thus, it must deal with the camera’s coverage region (Kaur and Singh 2014). Next, the overlapping of plant parts over the other regions causes complexity over the prediction of weed region from crops. Some pattern recognition approaches and algorithms are adopted for discriminating the weeds and crops (Matloob et al. 2015). It includes SVM, decision tree, artificial neural networks (ANNs), statistical pattern recognition, fuzzy logic, and so on for 2D vision processing.

Mubeen et al. (2014) models novel machine vision approaches using the stereo camera for differentiating soil, weeds, plants, and crops acquired from the cultivation fields under certain illumination conditions in the preliminary plant growth stage. The anticipated model considers normalised green modification, ANN, median filter application, morphological feature extraction, statistical threshold estimation, and image segmentation. Based on the observation, the author predicts the detection rate of weed with an accuracy of 90% (Ramprakash et al. 2013). However, there are some complexities identified in the generation of proper segmentation curves. This work adopts deep learning-based approaches to enhance the image quality and the prediction rate. The researcher’s interest turns towards this learning technique due to its significance of handling an extensive amount of data simultaneously and efficiently reduces the computation complexity (Rao and Chauhan 2015). Generally, deep learning approaches are used in various interesting applications on text generation, object classification, automatic machine translation, image colorisation, automatic game playing, medical applications, etc. Here, it is used for predicting and discriminating the weeds from the crop region (Rao and Ladha 2013; Rashid et al. 2012).

The significant target of this work is to model a novel deep learning-based model for discriminating the weeds from the paddy cultivation region using Regional Convolutional Neural Networks (R-CNN) to segment the weed objects from the input images. The process of segmentation improves prediction accuracy. For performing these processes, the images are acquired from the real-time environment, and it is collected for every stage of crop and weed growth. The segmented objects are measured with bounding box candidates. The network model extracts the ROI (region of interest) and evaluates the features to examine the bounding box coordinates and the object classes. The experimentation is performed in MATLAB 2016b environment. The outcomes show the significance of the anticipated R-CNN model compared with existing approaches like standard CNN, Support Vector Machine (SVM), Random Forest (RF), and multiple classifier model 1 and model 2, respectively. The simulation metrics include accuracy, precision, recall, F-measure, Matthews’s correlation coefficient, Mean Absolute Error, and Kappa coefficient.

The reaming part of the work is subdivided as follows: In Section 2, an extensive analysis regarding the weed detection approaches is discussed with its pros and cons. In Section 3, the anticipated R-CNN model is explained with data collection, segmentation, and experimentation process. Section 4 gives the numerical outcomes of the proposed R-CNN model with the elaborate discussion of numerical values with other approaches. Section 5 is a conclusion with future research ideas.

**Related works**

Various weed detection approaches, which are modelled to enhance the accuracy and weed detection, conflict mitigating the targets on providing weed controlling mechanisms for helping the farmers are discussed in this section. An extensive survey is conducted, and the
detection approaches are analyzed. Bakhshipour et al. (2017) discuss various biological morphology, spatial features, and contexts utilised to discriminate crops and weeds. These sorts of features are extracted by conducting various experiments and observations and modelling them towards image processing approaches. Berlin et al. (2021) performed wavelet texture features for validating the potential of differentiating the weed from the beet sugar crop. It is utilised to establish the texture features for all images and fed to ANN. Some co-occurrence texture features are described for multi-resolution image produced using wavelet transform. Finally, NN is used for labelling the classes as crops or weeds. The outcomes demonstrate that the wavelet texture features efficiently classify weeds from the crop region.

Rumpf et al. (2012) used threshold-based segmentation process to perform row-wise detection. The crop and weed were classified by 3D vector compression of the image using PCA. Sarker and Kim (2019) performed sequential SVM for classifying grains using shape parameters. Initially, sub-groups like crop plants, dicotyledonous weeds, and monocotyledonous weeds are classified. These species are distinguished from the provided group. Some features are chosen using SVM-based weighted and filters. Yu et al. (2019) discussed the emerging approaches for weed detection using learning approaches for directly extracting features and categorise crops or weeds based on automatic feature extraction. dos Santos Ferreira et al. (2017) used CNN for weed detection from soybean images. The model classifies and detects the weed among the leaf and grass. Here, the author achieves 90% accuracy for detecting grass and leaf weeds without soybean and soil in the background – (Wendel and Underwood 2016) modelled a self-supervised approach to differentiate weed using manual labelling. The author gathers the trained data for generating a self-supervised classification model to discuss the crop variation devoid of producing a manual dataset. The experimental outcomes demonstrate that the supervised model helps to discriminate crop from weed and manually label the training data. Kamalraj et al. (2021) adopts local features like scale-invariant and affine-invariant models for leaf image prediction. Here, the SVM classifier is merged with edge shape and surface colour for improving the entire prediction accuracy. Dyrman et al. (2017) developed a model for automatic weed detection using the presence of leaf occlusion. The network model is used for training and validating 17,000 weed annotations of wheat images. The model predicts 47% of weeds from the overlapped wheat. When some weeds, grasses, and other plants are exposed to some degree of overlap, the performance of the anticipated model is lesser. Thus, the model identifies some complexity during the creation of bounding boxes for the entire image.

Myers et al. (2015) discussed the discrimination of weeds from aerial images, and some drawbacks are identified with the adoption of flight altitude, cameras, and temporal resolution. Lottes et al. (2018) used GLCM and haralick descriptors for textural classification, and color-based prediction is used for normalised difference index. The experimental outcomes demonstrate that weed detection achieves 99% precision over the testing data and image resolution with the constrained accuracy. Potena et al. (2016) developed unmanned aerial vehicles with the multi-spectral camera and RGB for capturing images from the sunflower field. The overlapped images are used for generating high spatial resolution images. The image analysis is performed for predicting the weeds, crop rows, mapping soil culture, and soil mapping. Ronneberger et al. (2015) discuss weed detection with morphological and shapes from the reflectance spectra utilised for differentiating the model characteristics. These spectra gather information from the electromagnetic spectrum, and utilised to predict materials and objects. Some investigators use reflectance spectra for predicting weeds. Neelakandan and Paulraj (2020) used NN and SVM for detecting weeds from the corn cultivation region with hyperspectral reflectance. This investigation classifies and predicts weed management policies and nitrogen rates. The classification accuracy is 69% and 58%, respectively. Author discusses a model for differentiating weeds and crops based on spectral reflectance. The detection process relies on NN classifier. The prediction accuracy ranges from 30% to 95% based on the training samples fed to the classifier.

Guo et al. (2018) constructed a model for getting precise and accurate outcomes of weed prediction with the CNN classifier model and soil type to eradicate generalisation impact. The author evaluates the vegetation indices and specifications used to feed the input CNN. It provides accurate results even with the huge dataset. But, these sorts of indices, the anticipated model, give lesser prediction accuracy. Then, the encoder-decoder-based semantic segmentation is used for predicting the pixel-wise images with higher speed than the conventional CNN models. The author adopts the CNN model to categorise weed for real-time applications. It anticipates data reduction model and automatically selects images and specifies the informative samples as data summarisation. The summarised data is fed to CNN to attain accuracy over the original dataset. Table 1 depicts the comparison of various weed detection approaches.
The author adopts semantic segmentation for mapping and weed detection. DL-based semantic segmentation is composed of two preliminary blocks like decoding and encoding. The encoding blocks are considered with down sampling block that helps to extract features from the images, while the decoding process blocks feature space to image dimensionality. This segmentation process includes SegNet and UNet for a fully convolutional network over the top layers instead of the fully-connected layer. Ma et al. (2019) adopt CNN for image segmentation and classification due to its competency to automatically haul out the features. The CNN architecture is composed of max-pooling and convolutional layers. The former layer reduces the feature space dimensionality, and the latter model extracts the features. Followed by these layers are fully connected and top layers with softmax as the activation function for the multi-class classification process. The layers are added to improve the learning of complex features, and the model complexity is increased with more computational power. Various architectures like ResNet-50, VGG16, Xception, and Inception V3 are used for this process. However, these models produce computational complexity, and the bounding box coordinates of the images are not segmented properly. Thus, the proposed research concentrates on modelling the Regional Convolutional Neural Networks (R-CNN) to segment the weed objects from the input images. The objects (weeds) are measured with the bounding box candidates. The network model extracts ROI and analyses the features to infer the bounding box coordinates and the object classes.

**Methodology**

Our proposed weed predictor model (WPM) is composed of three essential parts: dataset acquisition, pre-processing, and segmentation. With the general CNN model, the feature extraction is performed with deep convolutional layers utilised to extract the features from the provided input images. R-CNN model is specifically designed for determining the bounding box coordinates for the normalised images. The overall model framework is shown in Figure 1.

**Data acquisitions**

The images used for this experimentation are real-time images that are collected from the paddy field. The crops are grown in the Thanjavur region, and the images are captured using the Canon SX730HS 20.3MP digital camera. Around 1000 images are recorded from the crop region, where 600 samples are utilised for training, and 400 samples are used for testing purposes. Thus, the validation ratio is partitioned as 60:40. Then, the images were gathered in the month of December 2020 as it is the specific season for crop cultivation (Figure 2).

**Pre-processing**

The images are acquired using the digital camera, the quality of the images obtained through digital camera is not so convincing. The images captured at different angles, and there were changes in illumination. Therefore, to fix this issue, image normalisation was performed. Here, dataset normalisation is performed over numeric features like missing pixel values. The non-numeric features are converted to numeric features with diverse attributes i.e. (0, 0, 1), (0, 1, 0), (1, 0, 0) over binary-vectors. Normalisation is depicted as the difference between the maximum and minimum values. The feature mapping is normalised with max-min normalisation and mathematically expressed as in Equation (1):

\[
\begin{align*}
    x_i &= \frac{x_i - \min}{\max - \min} \\
    \text{from Equation (1), } x_i &\text{ specifies data points, max specifies maximal values from data points for all features, and min specifies minimal data points. The features from the initial step of the model are extracted,}
\end{align*}
\]

Table 1. Comparison of various weed detection approaches.

| Parameters       | Detection mechanism | Accurate weed detection | Descriptions                              |
|------------------|---------------------|-------------------------|-------------------------------------------|
| Input            | Total RGB images    | Total RGB images        | Generalisation is used to construct the   |
|                  | (smaller dataset)   | (larger dataset)        | dataset and fulfils the target objects    |
|                  | SegNet              | ConvNet                 | using the huge dataset.                   |
| Classifier       | CNN                 | Generalisation          |                                      |
| Approaches       | Vegetation indices  | Summarisation with      |                                      |
|                  | and generalisation  | unsupervised data       |                                      |
|                  | impact (RGB images) | automatically selects   |                                      |
|                  |                     | the image from the      |                                      |
|                  |                     | dataset, which specifies|
|                  |                     | the informative images. |
|                  | Experimentation     | Accuracy ranges         |                                      |
|                  | Higher and faster   | from 95% and 97%        |                                      |
| Constraints      | accuracy            |                          |                                      |
|                  | Complex generalisation when images are less informative | The occurrence of pixel-wise images are failed, and the outcomes give lesser accuracy |
|                  |                     | The availability of the dataset shows some bottleneck approaches |
and it is performed using the convolutional and pooling layers. These modules produce feature maps for performing segmentation tasks.

**R-CNN segmentation**

Usually, Deep convolutional layers are proved to be more efficient for extraction of feature representation for the proved input images and efficiently perform segmentation and object detection. The pooling layers generate lower resolution features which cause loss of semantic information of the images. Thus, the convolutional layer plays a major part in addressing these issues and performs the region segmentation process. Here, region-based convolution is performed to enhance the bounding box detection accuracy. It is mathematically expressed as in Equation (2):

$$y[i] = \sum_{j=1}^{k} x[i + r.k]w[k]$$  

(2)

Here ‘y’ specifies the convolution layer output, ‘x’ specifies the input, ‘w’ specifies the filter weights and
'r' specifies the convolution rate for determining the feature extracted from the convolutional layers. The convolution rate is set as 1 which is similar to that of the convolution layers. The proposed R-CNN model is used for object detection and identifies the objects’ bounding boxes over the images. It is utilised to predict the weed over the paddy image. Moreover, the regional convolution model works as an end-to-end framework to merge the knowledge framework. The feature extraction module related to the input image is fed to the regional layer, which provides an output of the number of candidate regional factors. There is some Metrics to specify the probability of the box holding the influencing features. To extract the segmented region, anchors are initiated over the bounding box references. During feature mapping, anchors are proposed (3 aspects ratio * 3 scales) some authors are generated using the sliding windows. Here the candidate bounding box set is produced using fully convolutional layers. The numbers of anchors are the same as the number of candidate bounding boxes. Every candidate bounding box is provided with a score that specifies whether the boxes include the weed objects. It is observed that the fully convolutional network’s output does not directly possess the bounding box coordinates. However, the encoded forms of the bounding box coordinates are expressed as in Equation (3):  

\[
\begin{align*}
    t_x &= \frac{x - x_o}{w_o}; \\
    t_y &= \frac{y - y_o}{h_o}; \\
    t_w &= \log\left(\frac{w}{w_o}\right); \\
    t_h &= \log\left(\frac{h}{h_o}\right)
\end{align*}
\]  

(3)

Here \(x, y, w, h\) specifies the central region of bounding box coordinates, height, width, respectively. \(x_o, y_o, w_o, h_o\) specifies centre coordinates of anchors, width, and height, respectively. This work intends to find the central region of the bounding boxes. Thus, smaller amount of fully convolutional network outputs was chosen as the central region of the objects. Also, the selection operation diminishes the computational cost.

The regional CNN produces certain candidate bounding boxes. Here Region of Interest (ROI) pooling is essential. It helps to crop the smaller regions of feature mapping based on candidate bounding box coordinates. The ROI pooling output is given as the central region of interest with attention mechanisms. After cropping, the feature mapping sizes are the same. The process of extracting the ROI region is given: (1) the candidate bounding box coordinates are noticed over the image level. Therefore, the feature mapping size is transformed by dividing the model ratio. The pooling region is partitioned into \(k \times k\) blocks where ‘k’ specifies the output size. The maximum pooling techniques are used over smaller blocks, and the maximal values are pooled. Therefore, the pooled features are provided with \(k \times k\) size. The convolutional layer produces the encoded form of bounding box coordinates and the target object probability. When the number of regions is specified as ‘N’, the deep convolutional layer possesses a convolutional score and encoded coordinates. The bounding box prediction is computed based on the relationship among the encoded and candidate bounding box coordinates. The global and the local information of the weed images are combined with the corresponding regions of the paddy over diverse stages. When the feature mapping is performed over the large sizes, then the image holds local information; similarly, when the feature mapping is performed over the small-sized regions, then the images hold global information. Thus, the combinations of small and large feature maps are essential in regional segmentation and object detection. The input image size is cropped, and \(1 \times 1\) convolution is applied to the feature map (cropped), then based on concatenation with the cropped regions for the successive stages. The final output is the input of the regional CNN model. The anticipated regional-CNN model is trained, and the loss function is composed of the CNN intermediate layers. It is mathematically expressed as in Equation (4):

\[
f_{loss} = \text{loss}_{\text{intermediate\ layers}}
\]  

(4)

During training, the R-CNN model requires binary labels and specifies whether the image includes weed or not. The binary label is allocated based on the overlap of the weed over the paddy. When the anchor with the higher overlapped threshold, then it is labelled as positive. Here, the threshold is set as 0.6. When the overlapped value is lesser than 0.6, then the label is negative. The encoded region of the bounding box coordinates with ground truth values \(t_{x}, t_{y}, t_{w}, t_{h}\) is evaluated. The loss is evaluated among the ground truth values. It is mathematically expressed as in Equation (5):

\[
\text{Region} = \begin{cases} 
0.5 x^2 & \text{if } |x| < 1 \\
|x| - 0.5 & \text{else} 
\end{cases}
\]  

(5)

\[
R_{CNN} = \frac{1}{\text{class norm}} \sum_i L_{\text{binary loss}} (p_i, P_i) + \frac{1}{\text{region norm}} \sum_i p_i \cdot \text{Region} (t_i - t_i^*)
\]  

(6)

Here ‘i’ specifies the anchor index, \(p_i\) is prediction probability, \(P_i\) is the binary label, \(t_i\) is vector coordinates of the ith bounding box \(t_{x}, t_{y}, t_{w}, t_{h}\); \(t_i^*\) is ground truth values. \(L_{\text{binary loss}}\) is binary loss, ‘class norm’ and
'region norm' is region normalisation. The loss over the classifier is represented as 'L'. For classifier training, ROI is allocated with binary labels based on the overlapping of the weeds over the paddy with ground truth values. Here, the threshold is empirically set as 0.5. The bounding box regressions are evaluated using Equations (3) and (4). It relies on the candidate bounding box attained from the region-CNN model.

The proposed R-CNN-based segmentation outputs the coordinate bounding boxes, and the framework is provided with pixels points like \((x_0, y_0)\) and \((x_1, y_1)\). Based on these points, the central region of the image is provided as \((x_c, y_c)\), and it is described as in Equations (7) to (9):

\[
x_c = \frac{x_0 + x_1}{2} \quad (7)
\]
\[
y_c = \frac{y_0 + y_1}{2} \quad (8)
\]
\[
a = \frac{x_1 - x_0}{2} \quad (9)
\]
\[
b = \frac{y_1 - y_0}{2} \quad (10)
\]

Thus, the segmented outcomes of the R-CNN model are expressed in a standard form. The anticipated model works efficiently in segmenting the regions with better prediction accuracy over real-time data.

**R-CNN configuration**

The given dataset size 'N' consists of images \(I = \{i_1, \ldots, i_N\}\) with bounding boxes \(B = \{B_1, \ldots, B_N\}\), where the network configuration is done with the object segmentation. Here, R-CNN is employed with the parameters to categorise the central region over the images, i.e. both foreground and background. The proposed R-CNN model is simple and understandable for configuration. It is made of convolutional layers (two), ReLU, and max-pooling layers. The convolutional layer reduces the output layer by half the kernel size; the out padding with zero is performed to maintain the input size (See Table 2). The convolutional and the pooling layers are attached with dense neuron layers, and the output is related to both the background and the foreground images. Dropouts are added to the output and the dense layer to eliminate the over-fitting issues to perform regularisation. To eliminate the class imbalance issue, samples of equal number with \(K = 10^5\) patches from the training data for all classes, i.e. both foreground and background images. The weights are initialised with Gaussian distribution and adaptive gradient descent with learning rate \(\eta = 0.0015\) for a fixed number of epochs. Adaptive learning gives more benefits over the learning rate for all the features. Thus, it provides faster convergence and robust nature. The loss function is evaluated with cross-entropy among the encoding factors. During R-CNN training, the target regions are updated for all iterations. Data augmentation is essential for the training set (learned features) and helps eliminate over-fitting issues. The training data variations are increased with the Gaussian distribution intensity. The iterative nature of the R-CNN model provides global optima but is constrained towards the local optima. Thus, the resulting region-based segmentation relies on the initial regions. The segmentation process is nearer to the objects, and the bounding box coordinates with better accuracy and attains superior segmentation outcomes.

**Numerical results and discussion**

Here, the experimentation of the proposed R-CNN model is performed using a MATLAB simulation environment. The evaluation of R-CNN is done with various prevailing approaches like standard CNN, SVM, RF, MCS-1, and MCS-2, respectively. To evaluate the performance of the R-CNN, some metrics are considered. They are Accuracy, precision, recall, F-measure, MCC, MAE, and Kappa coefficient. The definitions for these metrics are given below:

1. **Accuracy**: specifies the appropriate proportion of total records in the given testing set. It is mathematically expressed as in Equation (11):

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)
\]

2. **Precision**: specifies the proportion of properly predicted weed to the predicted weed over the
testing process. It is mathematically expressed as in Equation (12):

$$
\text{Precision} = \frac{TP}{TP + FP}
$$

(12)

(3) **Recall**: specifies the proportion of properly predicted weed to the total weed samples over the testing set. It is mathematically expressed as in Equation (13):

$$
\text{Recall} = \frac{TP}{TP + FN}
$$

(13)

(4) **F-measure**: It is the measure of both recall and precision as shown in Equation (14):

$$
F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
$$

(14)

It is defined as True Positive (TP) – sample weed classified appropriately as a weed; True Negative (TN): standard samples classified appropriately as normal; False Positive (FP): normal sample classified wrongly as a weed; and False Negative (FN): sample paddy instances classified wrongly as weeds.

(5) **MCC**: Matthews’ correlation coefficient or phi coefficient is utilised to measure the binary classification quality. For instance, two classes- either ‘0’ or ‘1’ respectively. It is mathematically expressed as in Equation (15):

$$
\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}
$$

(15)

(6) **Kappa co-efficient**: It is used to evaluate the inter-rate reliability for the qualitative terms. It is considered a robust measure. It is expressed as in Equation (16):

$$
k = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}
$$

(16)

Here, Pr(a) is the measured probability; Pr(e) is the probability rate, either yes or no by random agreement.

(7) **Mean Absolute Error (MAE)**: It is the measure of errors among the paired observations specifying the phenomenon. It is expressed as in Equation (17):

$$
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i|
$$

(17)

Here y<sub>i</sub> is prediction, x<sub>i</sub> is the true value, and ‘n’ is the total number of data points.

The deep convolutional layers are used for feature learning are initialised using pre-trained model parameters. During training, the model is optimised, and the learning rate is initiated from 0.001 and drops to 0.0001 after successive iterations. The bounding box with maximal score is chosen to identify the image objects and the prediction accuracy. With the provided crop and weed-based images, the proposed R-CNN framework provides various bounding boxes. The model’s performance is compared with other approaches and shows the advantages of modelling the segmentation tasks as bounding box detection. With this process, the original segmentation problem is addressed efficiently by the proposed R-CNN model. Some other post-processing approaches are also required for the connected regions and handling the over-fitting and under-fitting issues. The outcomes enable that the outcomes of the models show stronger influence in a positive manner. The cropped feature maps from the feature extraction modules are modelled with R-CNN model and the computational complexity is also reduced considerably.

The evaluation of bounding boxes with direct learning helps to update the proposed R-CNN model outcomes in the improvements of accuracy. The bounding box is used for initialising the R-CNN method,

**Pseudo code 1: R-CNN**

**Input:**
Train the input dataset, i.e. train<sub>x</sub>; train<sub>y</sub>; test<sub>x</sub>; and test<sub>y</sub>.

**Output:** Regional-CNN weight and bias; Region<sub>smooth</sub> (x);

**Essential parameters:**
Bounding boxes $B = \{b_1, \ldots, b_n\}$; Learning rate of R-CNN = 0.001; Target_error rate; Loss function $l_{loss}$;

**Initialisation work:**
Set weight and bias of R-CNN;
Bounding box coordinates using y(i);

**Begin:**
1: Set essential parameters and perform initialisation work; // zero centre normalisation, padding, cross-entropy (network configuration);
2: Encode the bounding box coordinates using Equations (3) and (4);
3: while $l_{loss}$ = $l_{smooth}$intermediate layers;
4: for all training set;
5: train<sub>x</sub> and train<sub>y</sub> (label prediction);
6: end for
7: Evaluate smoothing region (x) using Equation (5);
8: Evaluate binary loss $l_{binary}$ with ‘class norm’ and ‘region norm’;
9: Measure central region using pixel points (x<sub>0</sub>, y<sub>0</sub>) and (x<sub>1</sub>, y<sub>1</sub>) ;
10: Attain segmented output from R-CNN using Equations (7) to (10);
11: end while; 12: End

and segmentation is visually nearer to the fully supervised R-CNN model. The R-CNN model facilitates pixel-wise segmentation from the image database with candidate bounding boxes. The proposed R-CNN model yields better accuracy by employing bounding boxes and allows data-driven feature extraction to enhance robustness. The iterative CNN model is...
provided with region analysis and the foreground region to update the score of target regions.

Moreover, it does not perform regularisation while updating the targets. It may not cause many issues during the segmentation of natural images, and a smoothing region is required during noisy image analysis. The regular shape of the input image is finely approximated with the candidate bounding boxes and reduces the false-positive rates during R-CNN training. The image contrast of the background is higher, and the intensity differences among the adjacent background of the objects are bigger than the conventional CNN model. The increase in segmentation accuracy is achieved efficiently with bounding box coordinates. The accuracy improves with the number of epochs, and the R-CNN models converge to various optima. The potential enhancement is identified during the target updating with the frequency of occurrence. The higher precision value with the learning-based models is underlined with superior performance compared to image segmentation approaches. The object-based learning from the collected images is favourable for model fitting. The R-CNN model is adjusted to cover the wider appearance variations. The objects with large class similarities are based on the appearance of the images. Generally, standard CNN architecture is used for image segmentation problems.

Table 3 depicts the comparison of various performance metrics like accuracy, precision, recall, and F-measure. The evaluation is done among the standard CNN, SVM, RF, MCS-1, and MCS-2, respectively. The accuracy of R-CNN is 83.3333% which are 10.4133%, 29.1633%, 27.0833%, 22.9133%, and 18.7533% higher than standard CNN, SVM, RF, MCS-1 and MCS-2. The precision of R-CNN is 83.5664% which is 4.5564%, 7.4764%, 7.4764%, 6.8964%, 5.6564%, and 4.2964% higher than other models. The recall of R-CNN is 83.333 which are 10.4133%, 29.1633%, 27.0833%, 22.9133%, and 18.7533% higher than standard CNN, SVM, RF, MCS-1 and MCS-2. F-measure of R-CNN is 83.3043% which is 11.8843%, 41.3243%, 37.4143%, 30.2443%, and 23.8043% higher than other models (See Figures 3 and 4). Table 4 depicts the comparison of MCC, MAE, and Kappa coefficient of R-CNN, SVM, RF, MCS-1, and MCS-2, respectively. The MCC value of R-CNN is 0.6690 which is 0.1533, 0.4605, 0.4108, 0.328, and 0.2558 higher than other models. The mean absolute error of R-CNN is 0.1667, which is comparatively lesser than other models, i.e. 0.1041, 0.2916, 0.2708, 0.2291, and 0.1875. The kappa coefficient of R-CNN is 0.6660,
which is 0.2077, 0.5872, 0.541, 0.4577, and 0.3743 higher than other models (See Figures 5 and 6). From the analysis, it is observed that the significance of the proposed model is comparatively higher than other approaches.

Figure 7 shows the samples of input images. Column 1 and Column 3 show the normal input image, and Column 2 and Column 4 shows the segmented image.

**Table 4.** Comparison of MCC, MAE, and Kappa of R-CNN model with existing approaches.

| Approaches   | MCC  | MAE  | Kappa coefficient |
|--------------|------|------|-------------------|
| Proposed R-CNN | 0.6690 | 0.1667 | 0.6660            |
| CNN          | 0.5157 | 0.2708 | 0.4583            |
| SVM          | 0.2085 | 0.4583 | 0.0833            |
| RF           | 0.2582 | 0.4375 | 0.1250            |
| MCS-1        | 0.3410 | 0.3958 | 0.2083            |
| MCS-2        | 0.4132 | 0.3542 | 0.2917            |

**Figure 4.** Recall and F-measure computation.

**Figure 5.** MCC and MAE computation.
Figure 6. Kappa co-efficient computation.

Figure 7. column 1 and column 3 (a), (c), (e), (g), (i), (k) are the Input image from the paddy field near Thanjavur district, column 2 and column 4 (b), (d), (f), (h), (j), (l) are the segmented image using the R-CNN model.
using the R-CNN model. Figure 8 depicts the discrimination of paddy and weed. The anticipated R-CNN model shows a better prediction rate when compared to other approaches.

**Results and discussion**

The vast advancements in deep learning approaches pave the way for achieving better accuracy in various real-time applications. This approach was first to create digital images of paddy crops and weeds from paddy fields using digital cameras, was fixed at different heights from the ground to make the method device-independent the soil and water background was removed. Texture, colour, and shape features were extracted. For this experimentation, the images were captured during the cultivation period (December – Feb) in the Thanjavur region. This proposed work intends to segment the weeds from the crop over the densely cultivated paddy crop. The unique model is proposed to address the segmentation problem by concurrent simulation of the task for object prediction. The experimentation performed shows the significant improvement in the detection accuracy with the proposed unique R-CNN model. The evaluation metrics were compared with other standard approaches like CNN, SVM, RF, MCS-1, and MCS-2. The accuracy of the proposed model was 83.33% which is superior than the other approaches.

**Authors’ contributions**

M. Vaidhehi – Methodology, Project Administration, Manuscript editing; C. Malathy – Software, Validation; Visualisation, Manuscript Review and editing.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

No funds, grants were received by any of the authors.

**Availability of data and material**

Not applicable.

**Code availability**

Not applicable.

**Notes on contributors**

*M. Vaidhehi* is an Assistant Professor at SRM Institute of Science and Technology, Kattankulathur campus Chennai, Tamilnadu, India. She completed her Bachelor’s degree and her Masters in Computer Science and Engineering from Anna University, Tiruchirappalli, TamilNadu, India in 2013. Now she is pursuing her PhD degree from SRM Institute of Science and Technology in computer science and technology, Chennai, Tamilnadu, India. Her research area is in Artificial Intelligent, Deep Learning and Machine Learning.

*C. Malathy* is a Professor in Department of Computer Science and Engineering, S.R.M Institute of Science and Technology, Kattankulathur campus, Chennai, India. She earned Ph.D. in Computer Science & Engineering from S.R.M Institute of Science and Technology, Chennai. She has more than Thirty years of experience in Teaching and Research. Her areas of interest are Image processing, Biometrics, Pattern recognition, Data Mining Computer architecture. She has published research papers in many international, national conferences and conferences and refereed journals.

**References**

Bakhshipour A, Jafari A, Nassiri SM, Zare D. 2017. Weed segmentation using texture features extracted from wavelet sub-images. Biosystems Eng. 157:1–12.

Berlin MA, Tripathi S, et al. 2021. IoT-based traffic prediction and traffic signal control system for smart city. Soft comput. doi:10.1007/s00500-021-05896-x.
Chauhan BS, Opena J. 2013. Weed management and grain yield of rice sown at low seeding rates in mechanized dry-seeded systems. Field Crops Res. 141:9–15.
Chauhan BS, Singh RG, Mahajan G. 2012. Ecology and management of weeds under conservation agriculture: a review. Crop Prot. 38:57–65.
Cheng X, Zhang Y, Chen Y, Wu Y, Yue Y. 2017. Pest identification via deep residual learning in complex background. Comput Electron Agric. 141:351–356.
dos Santos Ferreira A, Freitas DM, da Silva GG, Pistori H, Folhes MT. 2017. Weed detection in soybean crops using ConvNets. Comput Electron Agric. 143:314–324.
Dyrmann M, Jørgensen RN, Midtiby HS. 2017. Robo weed support-detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. Adv Ani Biosci. 8(2):842–847.
Gaire R, Dahal KR, Amgain LP. 2013. Effect of different mulching materials on weed dynamics and yield of direct seeded rice in Chitwan, Nepal. Agr J Nepal. 3:73–81.
Guo Y, Liu Y, Georgiou T, Lew MS. 2018. A review of semantic segmentation using deep neural networks. Int J Multimed Inf Retr. 7(2):87–93.
Joshi N, Singh VP, Dhyani VC, Chandra S, Guru SK. 2015. Weed management under different planting geometry in dry direct-seeded rice.
Kamalraj R, Ranjith Kumar M, Chandra Shekar Rao V, Anand R, Singh H. 2021. Interpretable filter based convolutional neural network (IF-CNN) for glucose prediction and classification using PD-SS algorithm. Measurement, 183, 2021, doi:10.1016/j.measurement.2021.109804.
Kaur S, Singh S. 2014. Influence of crop density on weeds, growth and yield of direct-seeded rice.
Khaliq A, Matloob A, Chauhan BS. 2014. Weed management in dry-seeded fine rice under varying row spacing in the rice-wheat system of Punjab, Pakistan. Plant Prod Sci. 17 (4):321–332.
Lottes P, Behley J, Milioto A, Stachniss C. 2018. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. IEEE Robotics and Automation Letters. 3(4):2870–2877.
Ma X, Deng X, Qi L, Jiang Y, Li H, Wang Y, Xing X. 2019. Fully convolutional network for rice seedling and weed image segmentation at the seeding stage in paddy fields. PloS one. 14(4):e0215676.
Matloob A, Khaliq A, Chauhan BS. 2015. Weeds of direct-seeded rice in Asia: problems and opportunities. Adv Agron. 130:291–336.
Mubeen K, Nadeem MA, Tanveer A, Jhala AJ. 2014. Effects of seeding time and weed control methods in direct seeded rice (Oryza sativa L.) J Ani Plant Sci. 24(2):534–542.
Myers D, Ross CM, Liu B. 2015. A review of unmanned aircraft system (UAS) applications for agriculture. 2015 ASABE Annual International Meeting (p. 1). American Society of Agricultural and Biological Engineers.
Neelakandan S, Pauraj D. 2020. A gradient boosted decision tree-based sentiment classification of twitter data. Int J Wavelets Multiresolution Inf Process. 18(4):205027, 1–21. doi:10.1142/S0219691320500277
Potena C, Nardi D, Pretto A. 2016. Fast and accurate crop and weed identification with summarized train sets for precision agriculture. International Conference on Intelligent Autonomous Systems, July. Springer, Cham; p. 105–121.
Ramprakash T, Madhavi M, Yakadri M. 2013. Influence of bispipyribac sodium on soil properties and persistence in soil, plant and grain in transplanted rice. Prog Res. 8(1):16–20.
Rao AN, Chauhan BS. 2015. Weeds and weed management in India-A Review.
Rao AN, Ladha JK. 2013. Economic weed management approaches for rice in Asia. The role of weed science in supporting food security by 2020. Proceedings of the 24th Asian-Pacific Weed Science Society Conference, Bandung, Indonesia, October 22–25, 2013. Weed Science Society of Indonesia. p. 500–509.
Rashid MH, Alam MM, Rao AN, Ladha JK. 2012. Comparative efficacy of pretilachlor and hand weeding in managing weeds and improving the productivity and net income of wet-seeded rice in Bangladesh. Field Crops Res. 128:17–26.
Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, October. Springer, Cham; p. 234–241.
Rumpf T, Römer C, Weis M, Sökefeld M, Gerhards R, Plümer L. 2017. Sequential support vector machine classification for small-grain weed species discrimination with special regard to Cirsium arvense and Galium aparine. Comput Electron Agric. 135:63–70.
Thenmozhi K, Reddy US. 2019. Crop pest classification based on deep convolutional neural network and transfer learning. Comput Electron Agric. 164:104906.
Wendel A, Underwood J. 2016. Self-supervised weed detection in vegetable crops using ground based hyperspectral imaging. 2016 IEEE international conference on robotics and automation (ICRA), May. IEEE; p. 5128–5135.
Yu J, Sharpe SM, Schumann AW, Boyd NS. 2019. Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. Pest Manage Sci. 75(8):2211–2218.