

# AI in Improving the Sustainability of Agricultural Crops

Subjects: Agricultural Engineering

Contributor: Ana Corceiro, Khadijeh Alibabaei, Eduardo Assunção, Pedro D. Gaspar, Nuno Pereira

The rapid growth of the world's population has put significant pressure on agriculture to meet the increasing demand for food. In this context, agriculture faces multiple challenges, one of which is weed management. While herbicides have traditionally been used to control weed growth, their excessive and random use can lead to environmental pollution and herbicide resistance. To address these challenges, in the agricultural industry, deep learning models have become a possible tool for decision-making by using massive amounts of information collected from smart farm sensors.

Keywords: weed detection ; deep learning ; weed classification ; support decision-making algorithm ; fruit detection ; disease detection ; CNN ; performance metrics ; agriculture

---

## 1. Disease Detection

Crop diseases reduce agriculture production and compromise global food security. A deep learning system that clearly identifies the specific timing and location of crop damage, leading to the spraying of herbicides only in affected areas can contribute to the moderation of resource use and environmental impacts <sup>[1]</sup>.

### Disease Detection in Individual Fruits

Afonso et al. <sup>[2]</sup> intended to categorize blackleg-diseased or healthy potato plants using deep learning techniques. An industrial RGB camera was employed to capture color pictures. Two deep convolutional neural networks (ResNet18 e ResNet50) were trained with RGB images with diseased and healthy plants. A model that had already been trained on the ImageNet dataset was employed to transfer learning for network weight initialization, and the Adam optimizer for weight optimization. Both networks were trained with a mini-batch size of 12 over a period of 100 epochs. The network ResNet18 was experimentally superior, with 94% of the images classified correctly. In contrast, only 82% of the ResNet50 classifications were correct. Precision was 85% and recall was 83% for the healthy class. The classifier used a rectified linear unit (RELU) activation to redefine the fully connected (FC) layer after linearly aggregating the output of the FC layer into a vector of size. The final network layer included a two-class linear classifier that enabled the binary classification utilizing logarithmic SoftMax activation (healthy versus blackleg).

Assunção et al. <sup>[3]</sup> presented a deep convolutional network to operate on mobile devices to categorize three peach disorders and healthy peach fruits (healthy, rot, mildew, and scab). In this research, the authors used transfer learning, data augmentation, and CNN MobileNetV2, which was trained on the ImageNet dataset to evaluate the outcomes of the disease classification in the comparatively small dataset of peach fruit disorders. The peach dataset was arranged with RGB images stored in the open website platform Forestry Images, Appizêzere, PlantVillage, University of Georgia, as well as from the Pacific Northwest Pest Management Handbooks, Utah State University. The ImageNet dataset was used to train the model initially (source task). Scab disease had the highest F1-score of 1.00, followed by the Rot and Mildew classes, each of which had a 0.96 F1-score. The classification for the Healthy class was an 0.94 F1-Score. The average F1-score for the model's overall performance was 0.96. No disease class was incorrectly classified by the model, which is crucial for disease study for control and infection. These successes highlight the promise of CNN for classifying fruit diseases with little training data. The model was also made to work with portable electronics.

According to Azgomi et al. <sup>[4]</sup>, a low-cost method was created for the diagnosis of apple disease in four different types, scab, bitter rot, black rot, and healthy fruits. The investigation employed a multi-layer perceptron (MLP) neural network. This technique was called Apple Diseases Detection Neural Network (ADD-NN). The images were captured with a digital camera. For picture clustering, the k-means technique was utilized in the study. Semi-automatic support vector machine (SVM) classification was carried out. After that, the disease was found by analyzing the attributes of the chosen clusters. A neural network was employed to enhance the procedure, make it completely automatic, and test the viability of increasing the created system's accuracy. Furthermore, the network was trained with the Levenberg–Marquardt algorithm. The accuracy of the procedure using various architectures for the neural network trained with 60% of the data was then

evaluated. The implementation of a two-layer formation with eight neurons in the first layer and eight in the second layer produced a maximum accuracy of 73.7%, according to the data.

### **Disease Detection in Areas of Crops**

In Kerkech et al. [5], the method proposed used a deep learning segmentation algorithm on UAV photos to identify the mildew disease in vines. The data was collected utilizing a UAV equipped with two MAPIR Survey2 camera sensors, comprising an infrared sensor and a RGB sensor configured for automated lighting. The SegNet architecture was used to divide visible and infrared pictures into four classes: symptomatic vine, ground, shadow, and healthy. When a symptom is seen in both the RGB and infrared pictures, it is considered that the disease has been discovered, and this was named "Fusion AND". In the second scenario, referred to as "fusion by the union", the symptom is declared identified if it is visible in either the infrared or RGB picture and is denoted by the sign "fusion OR". The model trained with RGB images outperformed the model trained with infrared images, with an accuracy of 85.13% and 78.72%, respectively. Moreover, the model fusion OR outpaced the fusion AND with an accuracy of 92.23% and 82.80%, in that order. For visible and infrared photos, SegNet's runtime on a UAV image was estimated to be 140 s. Less than 2 s are required for the merging of the two segmented pictures.

## **| 2. Weed Detection**

In addition to disease, weeds are seen as a common danger to food production. The technologies described in this study may be used to power weed-detecting and weed-eating robots [4].

### **Weed Detection in Individual Plants**

In accord with Sujaritha et al. [6], fuzzy real-time classifiers were used to find weeds in sugar cane fields. Using a Raspberry Pi microcontroller and appropriate input/output subsystems including two different cameras, motors with power supplies, and tiny light sources, a robotic prototype was created for weed detection. During the movement of the robot, a divergence in the established course might occur due to obstacles in the field. An automatic image classification system was constructed, and it used a fuzzy real-time classification method and extracted leaf textures.

Among nine distinct weed species, the proposed robot prototype accurately recognizes the sugarcane crop. With a processing time of 0.02 s, the system identified weeds with an accuracy of 92.9%.

Milioto et al. [7] developed a new methodology for crop-weed classification using data taken with a 4-channel RGB and NIR camera, which depends on a modified encoder-decoder CNN. Three separate inputs were used to train the networks: RGB, RGB and near-infrared (NIR) images, and 14 channels including vegetation indices RGB, Excess Green (ExG), Excess Red (ExR), Color Index of Vegetation Extraction (CIVE), and Normalized Difference Index (NDI). To supplement the CNN with additional inputs, the authors first computed various vegetation indices and alternative interpretations that are often employed in plant categorization.

The authors found that the model performed better when additional channels were added to the input to the CNN. The network using RGB was 15% quicker to converge to 95% of the final accuracy than the network using the NIR channel. In terms of object-wise performance, the model achieved an accuracy of 94.74%, a precision of 98.16% for weeds, and 95.09% for crops. For recall, the system accomplished 94.79% for weeds and 94.17% for crops. The intersection over the union was 80.8%.

Lottes et al. [8] designed a sequential model encoder-decoder FCN for weed identification in sugar beet fields. The dataset was collected using a field robot, namely, BoniRob, with a 4-channel RGB+NIR camera. The processing model used 3D convolution to analyze five images in a series, creating a sequence code that was then used to learn sequential information about the weeds in the five images in a series. With the help of an addition known as the sequential module, it was possible to use picture sequences to implicitly encode local geometry. Even if the optical appearance or development stage of the plant changes between training and test time, this combination improves generalization performance.

The results indicated that, in comparison to the encoder-decoder FCN, the encoder-decoder with a sequential model raised the module's F1-score by around 11 to 14%. The suggested model outperformed encoder-decoder FCN without a sequential model, with an F1-score of 92.3.

Ma et al. [9] proposed an image segmentation procedure with SegNet for rice seedlings and weeds at the seedling stage in the paddy field based on fully convolutional networks (FCN). The model was then compared with another model, namely, U-Net. In this study, RGB color images were captured in seedling rice paddy fields. SegNet was developed using

a symmetric structure for encoding and decoding, which was utilized to extract multiscale features and increase feature extraction accuracy. This AI method can directly extract the characteristics from the original RGB photos as well as categorize and identify the pixels in paddy field photographs that belong to the rice, background, and weeds. The primary goal of this study was to evaluate how well the suggested strategy performed in comparison to a U-Net model.

Ferreira et al. <sup>[10]</sup> analyze the performance of unsupervised deep clustering algorithms in real weeds datasets (Grass-Broadleaf dataset, and DeepWeeds), for the identification of weeds in a soybean field. Deep Clustering for Unsupervised Learning of Visual Features, and Joint Unsupervised Learning of Deep Representations and Image Clusters (JULE) are two contemporary unsupervised deep clustering techniques (DeepCluster).

The DeepCluster model was built using AlexNet and VGG16 as a baseline to obtain features, and K-means were implemented as the clustering algorithm.

Analyzing the two clustering algorithms evaluated, JULE performed more poorly than DeepCluster, in terms of the normalized mutual information (NMI), and accuracy. In JULE, for the first dataset, the results of MNI and ACC were 0.28% and 65.6%, respectively, for 80 clusters. In the second dataset, the results of MNI and ACC were 0.08% and 25.9%, respectively, for 160 clusters. On the other hand, in DeepCluster for the first dataset, the results of MNI and ACC were 0.41% and 87%, respectively, for 160 clusters. For the second dataset, the results of MNI and ACC were 0.26% and 51.6%, respectively, for 320 clusters.

In Wang et al. <sup>[11]</sup>, pixel-wise semantic segmentation of weed and crop was examined using an encoder-decoder deep learning network. The two datasets used in the study, specifically, sugar beet and oilseed, were collected under quite varied illumination conditions. Three picture improvement techniques, Histogram Equalization (HE), Auto Contrast, and Deep Photo Enhancer, were examined to lessen the impacts of the various lighting situations. To improve the input to the network, several input representations, including different color space transformations and color indices, were compared. The models were trained with YCrCb and YCgCb color spaces and vegetation indices such as NDI, NDVI, ExG, ExR, ExGR, CIVE, VEG, and MExG. The results demonstrated that while the inclusion of NIR information significantly increased segmentation accuracy, images without NIR information did not improve segmentation results, demonstrating the value of NIR for accurate segmentation in low light conditions. The segmentation results for weed detection obtained by applying deep networks and image enhancement techniques in this work were encouraging. The model trained using NIR pictures attained a mIoU of 87.13% for the sugar beetroot dataset. For the oilseeds' dataset, the models were trained with RGB images only, and outperformed the other models with a mIoU of 88.91%. The best accuracy was 96.12%.

Kamath et al. <sup>[12]</sup> applied semantic segmentation models, namely, UNet, PSPNet, and SegNet in paddy crops and two types of weeds. The paddy field image collection was compiled from RGB photographs from two separate sources using two digital cameras. Two datasets were then created; only weed plants were included in Dataset-1, whereas paddy crop and photos of weeds were included in Dataset-2. A segmentation architecture using the ResNet-50 base model was built in PSPNet. A feature map for PSPNet was produced from the base network. On these pooled feature maps, convolution was used before feature maps were upsampled and concatenated. The use of a final convolution layer results in segmented outputs. The encoder-decoder framework used by the UNet design was constructed using the ResNet-50 base model. This model used skip connections which are additional connections that join down sampling layers with up sampling layers. The rebuilding of segmentation boundaries with the aid of skip connection after down sampling results in a more accurate output image. The VGG16 network and the encoder network used by the SegNet model are topologically identical. Each encoder layer has a matching decoder layer, and then each pixel receives class probabilities from a multi-class SoftMax classifier.

Using the playment.io program, photos were annotated, and each pixel was labelled to a categorization from one of four categories: Background-0, Broadleaved weed-1, Sedges-2, and Paddy-3. PSPNet outperformed SegNet and UNet in terms of effectiveness. The mean IoU for PSPNet was 0.72 and, the frequency weighted IoU was 0.93, whereas for SegNet and UNet, the mIoU values were 0.82 and 0.60, respectively. Finally, the frequency weighted IoU values were 0.74 and 0.38, respectively.

Mu et al. <sup>[13]</sup> developed a project to identify weeds in photos of cropping regions using a network model based on Faster R-CNN. Beyond that, another model combining the first one with Feature Pyramid Network (FPN) was developed for improved recognition accuracy. Images from the V2 Plant Seedlings dataset were used; this file includes photos in different weather conditions. The Otsu technique was applied to transform the obtained greyscale pictures into binary images to segregate the plants. Clear photos of the plants were obtained after processing. The convolutional features are shared using the Faster R-CNN deep learning network model, and feature extraction is done by fusing the ResNeXt

network with FPN, to improve the model's weed identification detection accuracy. The experimental results show that the Faster R-CNN-FPN deep network model obtained greater recognition accuracy by employing the ResNeXt feature extraction network and combining it with the FPN network. Both models achieved good results; however, the prototype with FPN reached an accuracy of 95.61%, a recall of 87.26%, an F1-value of 91.24, an IoU of 93.7, and a detection time of 330ms. The model without FPN achieved the following results for the same metrics, 92.4%, 85.2%, 88.65%, 89.6%, and 319 ms.

Assunção et al. [14] explored the optimization of the weed-specific semantic segmentation model at model DeeplabV3 with a MobileNetV2 backbone, as well as its impacts on segmentation performance and inference time. In this study, the experiments were conducted with  $DM = 1.0$  and  $DM = 0.5$ . The OS hyperparameter is the ratio of the size of the encoder's final output feature map to the size of the input image. Values of 8, 16, and 32 were chosen for OS to explore the trade-off between accuracy and inference time since this hyperparameter affects segmentation accuracy and inference time. There are three sections to this piece. There were two datasets utilized in the first one. To train and test the models, the Crop Weed Field Image Dataset (CWFID) dataset, which includes crops (carrots) and weeds, was employed. The second section of the process utilized crop and weed photos for the model's training and validation. By choosing several model hyperparameters and using model quantization, the model was optimized both before and after training. The primary goal is to extract the characteristics of the input image.

To obtain the performance necessary for the application, the depth multiplier (DM) and output stride (OS) hyperparameters of the MobilinetV2 were modified (i.e., light weight and fast inference time). The checkpoint files were then transformed into a frozen graph using a TMG framework tool (script). Finally, using the TensorRT class converter, the frozen graph was modified (optimized) to operate on the Tensorflow Real-Time (TensorRT) engine.

The semantic segmentation model was utilized in the most recent test of the robotic orchard rover created by Veiros et al. (2022). In this study, the accuracy and viability of a computer-vision framework were evaluated using a system for spraying pesticide on weeds. A Raspberry Pi v2 camera module with an 8-megapixel Sony IMX219 sensor was used to take the video pictures. The actuators that the Jetson Nano device controls are the herbicide container, pressure motor, a DC motor that applies pressure to it, manipulator motor, a stepper motor that moves the axis of the Cartesian manipulator, nozzle relay, a relay that opens and closes the spray valve, and spray nozzle.

According to the study results of the second test, segmentation performance mean intersection over union (mIOU) declined by 14.7% when employing a model hyperparameter DM of 0.5 and the TensorRT framework compared to a DM of 1.0 and no TensorRT. The model with the best segmentation performance has a 75% mIOU for OS = 8 and DM = 1.0. The model with a DM of 0.5 and OS of 32 had the lowest performance, which was 64% mIOU.

### **Weed Detection in Areas of Crops**

Peña et al. [15] created a study to evaluate the effectiveness and constraints of remotely sensed imagery captured by visible and multispectral cameras in an unmanned aerial vehicle (UAV), for early weed seedling detection. The objectives of the work were: to choose the best sensor for enhancing vegetation (weed and crop) and bare soil class discrimination as affected by the vegetation index applied; to design and test an algorithm object-based image analysis (OBIA) technique for crop and weed patch detection; and to determine the best arrangement of the UAV flight for the altitude, the type of sensor (visible-light + near-infrared multispectral cameras vs. visible-light), and the date of flight.

The OBIA procedure combined object-based characteristics such as spectral values, position, and orientation, as well as hierarchical relationships between analysis levels. As a result, the system was designed to identify crop rows with high accuracy using a dynamic and self-adaptive classification process and to label plants outside of crop rows as weeds.

The maximum weed detection accuracy, up to 91%, was found in the color-infrared pictures taken at 40 m and on date 2 (50 days after seeding), when plants had 5–6 true leaves. The images taken earlier than date 2 performed significantly better than the ones taken subsequently at this flight level. With a higher flight altitude, the multispectral camera had superior accuracy, while the visible light camera had higher accuracy at lower altitudes. The errors are due to the higher altitudes as a consequence of the spectral mixture between bare soil elements and sunflowers that occurred at the perimeters of the crop-rows.

Huang et al. [16] used photos from a UAV Phantom 4 to create an accurate weed cover map in rice fields, to detect weeds and rice crops. The Fully Convolutional Network (FCN) approach was proposed for preparing a weed map of the captured images. In the training phase, the image-label pairings from the training set that correlates pixel-to-pixel are fed into the

FCN network. The network converts the input picture into an output image of the same size, and the output image is applied to calculate the loss as an objective function together with the ground truth label (GT label).

According to the investigational results, the performance of the FCN technology was very effective. The general accuracy of the system reached 0.935, its weed detection accuracy reached 0.883, and an IoU 0.752, indicating that this algorithm can provide specific weed cover maps for the UAV images under consideration.

Bah et al. <sup>[17]</sup> developed a completely automated learning technique for weed detection in bean and spinach fields using UAV photos utilizing ResNet18 with a selection of unsupervised training datasets. This algorithm created super-pixels based on k-mean clustering.

A simple linear iterative clustering (SLIC) algorithm was used to construct a marker and define the plant rows after the Hough transform was used to determine the rate of plant rows on the skeleton. Super pixels were produced by this technique using k-mean clustering.

Other models, namely, SVM and RF, were used to compare the model's performance. ResNet18 performs better overall than SVM and RF in supervised and unsupervised learning techniques. The model achieved an accuracy of 0.945 and a kappa coefficient of 0.912.

In line with Osorio et al. <sup>[18]</sup>, three different weed estimation methods were proposed based on deep learning image processing and multispectral images captured by a drone. An NDVI index was used in conjunction with these techniques. The first technique uses histograms of oriented gradients (HOG) as a feature descriptor and is based on SVM. The ground and other aspects that are unrelated to vegetation are covered by a mask that is created by NDVI. These objects' characteristics are retrieved using HOG and are then used as inputs by a support vector machine that has already been trained. The SVM determines whether the identified items fall into the lettuce class. The second approach employed a CNN based on YOLOv3 for object detection. An algorithm removes crop samples from the image using model's bounding box coordinates after it has been trained to recognize the crop. After that, a green filter binarizes the picture, turning the pixels that do not have any vegetation into black and the ones that the green filter accepts into white. Finally, vegetation that does not match the crop is highlighted, making it easier to calculate the percentage of weeds in each image. The last method was to apply masks on the CNN, to obtain an instance segmentation for each crop. RCNN extracts 2000 areas from the picture using the "selective search for object recognition" method. They feed data into the Inception V2 CNN in this case, and it extracts characteristics used by an SVM to categorize the item into the appropriate category. Centered on the metrics that were used, the F1-scores for crop detection using this approach were 88%, 94%, and 94%, respectively. The accuracy was 79%, 89%, and 89%. The sensitivity was 83%, 98%, and 91%. The specificity was 0%, 91%, and 98%. Finally, the precision was 95%, 91%, and 94%.

Considering the version of the YOLO model used, it is important to say that there are currently more up-to-date versions. YOLOv4 is an advanced real-time object detection model that was introduced as an improvement over the previous versions of YOLO. Developed by a team at the University of Washington, YOLOv4 boasts a significantly improved performance in terms of accuracy and speed compared to its predecessors. It includes a new architecture that incorporates spatial pyramid pooling and a backbone network based on CSPDarknet53. This architecture allows for more efficient use of computing resources, resulting in faster processing times and improved accuracy. Additionally, YOLOv4 uses a combination of anchor boxes and dynamic anchor assignment to improve object detection accuracy and reduce false positives. Another notable feature of YOLOv4 is its use of a modified loss function that includes a term to penalize incorrect classifications of small objects. This leads to better performance on small object detection tasks <sup>[19]</sup>.

YOLOv5 is a state-of-the-art object detection and image segmentation model introduced by Ultralytics in 2020. It builds on the success of previous YOLO models and introduces several new features and improvements. One of the key innovations in YOLOv5 is its use of a new, more efficient architecture based on a single stage detection pipeline. This pipeline uses a feature extraction network combined with a detection head, which allows for faster processing times and improved accuracy. Additionally, YOLOv5 introduces a range of new anchor-free object detection methods, including the use of center points, corner points, and grids <sup>[19]</sup>.

YOLOv8 is an advanced object detection and image segmentation model that was developed by Ultralytics. It is an improvement over previous YOLO versions and has gained popularity among computer vision researchers and practitioners due to its high accuracy, speed, and versatility. One of the main strengths of YOLOv8 is its speed, which enables it to process large datasets quickly. Additionally, its accuracy has been improved through a more optimized network architecture, a revised anchor box design, and a modified loss function. This results in fewer false positives and false negatives, leading to better overall performance. Overall, YOLOv8 is an excellent tool for computer vision

applications and offers many advantages over previous models. Its speed, accuracy, and versatility make it an ideal choice for a broad range of tasks, including object detection, image segmentation, and image classification [20].

Islam et al. [21] used three types of approaches, namely, KNN, RF, and SVM to detect weeds in crops. The images were acquired from an RGB camera coupled in a UAV, in an Australian chilli farm, and then pre-processed using image processing methods. Red, green, and blue bands' reflectance was extracted, and from there, the authors deduced vegetation indicators such as the normalized red band, normalized green band, and normalized blue band. The pre-processed pictures' features were extracted using MATLAB, which was also utilized to simulate machine learning-based methods. The experimental findings show that RF outperformed the other classifiers. In light of this, it is clear that RF and SVM are effective classifiers for weed detection in UAV photos. RF, KNN, and SVM each had accuracy results of 0.963, 0.628, and 0.94. Recall and specificity were 0.951 and 0.887, 0.621 and 0.819, and 0.911 and 0.890 with RF, KNN, and SVM, respectively. With RF, KNN, and SVM, respectively, the accuracy, false positive rate (FPR), and kappa coefficient were 0.949, 0.057, and 0.878; 0.624, 0.180, and 0.364; and 0.908, 0.08, and 0.825.

### 3. Weed Classification

Another crucial component of agricultural management is the categorization of species (such as insects, birds, and plants). The conventional human method of classifying species takes time and calls for subject-matter experts. Deep learning can analyze real-world data to provide quicker, more accurate solutions [1].

#### Weed Classification in Individual Plants

According to Dyrmann et al. [22], a convolutional neural network was developed to recognize plant species in color images. The images originated from six different datasets namely, Dyrmann and Christiansen (2014), Robo Weed Support (2015), Aarhus University—Department of Agroecology and SEGES (2015), Kim Andersen and Henrik Midtby, Sjøgaard (2005), Minervini, Scharr, Fischbach, and Tsafaris (2014). The six datasets include all pictures taken during lighting-controlled events and photographs taken on mobile devices while out in the field during varying lighting circumstances.

To identify green pixels, a straightforward excessive green segmentation was employed. After that, batch normalization makes sure that the inputs to layers always fall within the same range. The network's activation function (ReLU) adds non-linear decision boundaries. Max pooling is a procedure that shrinks a feature map's spatial extent and gives the network translation invariance. In this study, the network's layering was decided upon by assessing the network's filtering power and coverage.

The training was ended after 18 epochs to get the maximum accuracy feasible without over-fitting the network. With an average accuracy of 86.2%, the network's categorization accuracy varied from 33% to 98%. With accuracy rates of 98%, 98%, and 97%, respectively, Thale Cress (*A. thaliana*), Sugar Beet (*B. vulgaris*), and Barley (*H. vulgare* L.) were frequently accurately diagnosed. However, Broadleaved Grasses (*Poaceae*), Field Pansy (*Viola arvensis*), and Veronica (*Veronica*) were frequently misclassified. Just 46%, 33%, and 50% of these three species received the proper classification. Overall, the classes with the greatest number of species also had the greatest categorization accuracy. As a result, classes with fewer picture samples made a smaller total loss.

Andrea et al. [23] demonstrated the creation of an algorithm capable of classifying and segmenting images. It uses a convolutional neural network (CNN) to separate weeds from maize plants in real-time. This discrimination was performed using four types of CNN, namely, AlexNet, LeNet, sNet, and cNet. A multispectral camera was used to acquire RGB and NIR images for segmentation and classification. A dataset created during the segmentation phase was used to train the CNN. Each of the four CNN models was trained using the same dataset and solver of type Adam after being selected.

The most successful algorithms offer great potential for real-time autonomous systems for categorizing weeds and plants. The network that produced the best results was the cNET of 16 filters. It had a training accuracy of 97.23% and used a dataset of 44,580 segmented pictures from both classes.

Gao et al. [24] proposed a hyperspectral NIR snapshot camera for classifying weeds and maize by measuring the spectral reflectance of an interest zone (ROI). The aim of this work was to identify the relevant spectral wavelengths and key features for classification, investigate the viability of weed and maize classification using a near infrared (NIR) snapshot mosaic hyperspectral camera, and provide the best parameters for a random forest (RF) model construction. In that work, 185 features were retrieved using vegetation indices (VIs), specifically, NDVI and RVI.

According to the findings, the ideal random forest model with 30 crucial spectral properties can successfully identify the weeds *Convolvulus arvensis*, *Rumex*, and *Cirsium arvense*, as well as the crops *Zea mays*. It was demonstrated that *Z. mays* can be identified with 100% recall (sensitivity) and 94% precision (positive predictive values). The model accomplished precision and F1 scores of 0.940 and 0.969, 0.959 and 0.866, 0.703 and 0.697, and 0.659 and 0.698, for crop *Zea mays* and weeds *Convolvulus arvensis*, *Rumex* and *Cirsium arvense*, respectively.

Bakhshipour and Jafari [25], using shape characteristics, utilized a Support Vector Machine (SVM) and an artificial neural network (ANN) classifier to categorize four different species of weeds and a sugar beet crop. Pictures were captured by using a weed robot with a camera, providing RGB images. Multi-layer feed-forward perceptron ANN was created using the Levenberg–Marquardt (LM) back-propagation learning method and two hidden layers. Principal Component Analysis (PCA) was employed as a feature selection method to reduce the initial 31 feature expressions into four components. The PCA values were then employed in SVM.

Both ANN and SVM correctly classified the sugar plants, with an accuracy of 93.33% and 96.67%, respectively. Compared to the sugar beet crop, the weeds were correctly identified by ANN and SVM 92.50% and 93.33% of the time, respectively. With an overall accuracy of 92.92% and 95%, respectively, both ANN and SVM were able to detect the shape-based patterns and categorize the weeds quite well.

Sa et al. [26] performed weed and sugar beet classification using a CNN with multispectral images collected by a MAV. These images were converted to SegNet format. The information gathered from this field was divided into photographs with only crops, pure weeds, or a combination of crops and weeds. For improved class balance, the frequency of appearance (FoA) for every single class is modified depending on the training dataset. With changing input channel sizes and training settings, the authors trained six distinct models, assessed them quantitatively using AUC and F1-scores as metrics, and then compared the results.

The learning rate for the training model was set to 0.001, the batch size was 6, the weight delay rate was 0.005, and the maximum iterations were 640 epochs. This model was able to achieve an average accuracy of 80% using the test data, with an average F1-score of 0.8. However, spatiotemporal inconsistencies were found in the model due to limitations in the training dataset.

Yang et al. [27] investigated deep learning techniques for hyperspectral image classification. The authors designed and developed four deep learning models: a two-dimensional CNN (2-D-CNN); a three-dimensional CNN (3-D-CNN); a region-based 2-dimensional CNN (R-2-D-CNN); a region-based 3-dimensional CNN (R-3-D-CNN). The objective was that a 2-D-CNN worked in the spatial context, while a 3-D-CNN worked in both spectral and spatial factors of the hyperspectral images retrieved from six datasets, viz., Botswana Scene, Indian Pines Scene, Salinas Scene, Pavia Center Scene, Kennedy Space Center, and Pavia University Scene.

The patch and feature extraction and the label identification steps make up the 2-D-CNN model. The primary distinction is that the 3-D-CNN model contains an additional reordering step. The D hyperspectral bands are rearranged in this phase in ascending order. A multiscale deep neural network is used by the R-2-D-CNN model to fuse numerous shrinking patches into multilevel instances, which are then used to make predictions. The primary distinction is that the 3-D-CNN model makes use of 3-D convolution operators whereas the R-2-D-CNN model do so use their 2-D equivalents.

An effective hyperspectral image classification process should consider both the spectral factor and the spatial factor since both have an impact on the class label prediction of a pixel. With this knowledge, the proposed deep learning models, namely the R-2-D-CNN and the R-3-D-CNN, achieved better results. The best results of the first network, in one of the datasets, were 99.67% and 99.89%, which correspond to values of average accuracy of each class (AA), and overall accuracy of all classes (OA), respectively. In the second model, the best results were 99.87% and 99.97%, for the same metrics.

Yashwanth et al. [28] implemented an image Classification System using the Deep Learning function. KERAS API in combination with the Tensorflow backend has been used in Python. Images of nine different crops and their respective weeds have been collected (*wheat-Parthenium*; *Soybean-Amaranthus Spinosus*; *Maize-Dactyloctenium Aegyptium*; *Brinjal-Datura Fatuosa*; *Castor-Portulaca Oleracea*; *Sunflower-Cyperus Rotundus*; *Sugarcane-Convolvulus Arvensis*; *Paddy-Chloris Barbata*; *Paddy-Echinochloa colona*). In the first stage, images that will be used to train the neural network are pre-processed. The input layer stores the image's pixels in the form of arrays. The "ReLU" activation function is used in the next step to obtain the image's corrected feature map. To accomplish edge detection, pooling is employed. The matrix gets flattened after using this flattened function. The thick layer receives this feeding. The object in the image is recognized by a completely linked layer.

The model was tested using nine different types of crops and the corresponding weeds, and the highest accuracy was found to be 96.3%. The provided photos were correctly categorized as either plants or weeds.

Jin et al. [29] created an algorithm for robotic weed eradication in vegetable farms based on deep learning and image processing. Images were captured in the field using a digital camera. Bounding boxes were drawn on the vegetable in the input photos as a manual annotation. In CenterNet, each item is represented by a single point, and object centers are predicted using a heatmap. Estimated centers are obtained from the heatmap's peak values using a Gaussian kernel and an FCN. Using a Gaussian kernel and focal loss, each ground truth key point is transformed into a smaller key-point heatmap to train the network. A color index was established and assessed using Genetic Algorithms (GAs) in accordance with Bayesian classification error to extract weeds from the background.

The trained CenterNet earned an F1-score of 0.953, an accuracy of 95.6%, and a recall of 95.0.

In El-Kenawy et al. [30], a new methodology based on metaheuristic optimization and machine learning was proposed, which aims to classify weeds based on wheat images acquired by a drone. Three models were proposed, specifically, artificial neural networks (NNs), support vector machines (SVMs), and the K-nearest neighbors' algorithm (KNN). The ANN was trained across a public dataset, through transfer learning and feature extraction. According to AlexNet, a binary optimizer is further suggested to improve the feature selection procedure and choose the optimal collection of features. A collection of assessment criteria is used to evaluate the efficacy of the feature selection algorithm to analyze the performance of the suggested technique. The suggested model used two more different types of machine learning models, namely, SVM and KNN, to improve the parameters. This classifier is improved by a brand-new optimization approach that combines grey wolf (GWO) and sine cosine optimizers (SCA). These suggested classifiers contribute to a creation of a hybrid algorithm.

The results demonstrate that the recommended technique works better than other alternatives and enhances classification accuracy, with a detection accuracy of 97.70%, an F1-score of 98.60%, a specificity of 95.20%, and a sensitivity of 98.40%.

Sunil et al. [31] analyzed the performance of a deep learning model for weed detection in photos with non-uniform and uniform backgrounds. Four Canon digital cameras were used to capture the weed and crop shots, namely, Palmer amaranth, horseweed, redroot pigweed, waterhemp, ragweed, and kochia, and crop species of sugar beets and canola. Weed classification models were developed using deep learning architectures, namely, Convolutional Neural Network (CNN) based on a Residual Network (ResNet50), and Visual Group Geometry (VGG16). The uniform background scenario data, non-uniform background scenario data, and combined-datasets scenarios created after combining both scenarios' data were trained using the ResNet50, and VGG16.

With an average f1-score of 82.75% and 75%, respectively, the VGG16 and ResNet50 models built from non-uniform backdrop pictures performed well on the uniform background. The performance of the VGG16 and ResNet50 models, which were built using uniform backdrop photos, did not fare as well, with average f1-scores of 77.5% and 68.4%, respectively, on non-uniform background images. The f1-score value of 92% to 99% was achieved by a model that was trained using fused information from two background circumstances.

Sunil et al. [32] compared the classification models of Support Vector Machine (SVM) and deep learning-based Visual Group Geometry 16 (VGG16) utilizing RGB picture texture information to categorize weeds and crop species. Six crop species as well as four weeds (horseweed, kochia, ragweed, and waterhemp) were classified using the SVM and VGG16 deep learning classifiers (the crop species were black bean, canola, corn, flax, soybean, and sugar beets). Gray-level co-occurrence matrix (GLCM) features and local binary pattern (LBP) features are two different categories of texture characteristics that were retrieved from the grayscale picture. After this, a machine learning classifier was built by operating a SVM and VGG16.

All SVM model classifiers have fallen short in comparison to the VGG16 model classifiers. The findings showed that the VGG16 model classifier's average F1 results varied from 93% to 97.5%, while the average F1-score results of SVM ranged from 83% to 94%. In the VGG16 Weeds-Corn classifier, the corn class achieved a F1-score value of 100%.

## **4. Fruit Detection**

Fruit quality detection is a technique for automatically evaluating the quality of fruits based on several aspects of a picture, such as color, size, texture, and form, among others. The main element preventing adverse health issues in people is fruit quality. In the food business and agriculture specifically, automatic detection is crucial.

## Fruit Detection in Individual Plants

Mao et al. [33] proposed a Real-Time Fruit Detection model (RTFD), a simple method for edge CPU devices that can identify fruit, specifically strawberries and tomatoes. The PicoDet-S model-based RTFD enhances the efficiency of real-time detection for edge CPU computing devices by enhancing the model's structure, loss function, and activation function. Two datasets were used with pictures taken in different conditions; the tomato dataset was compiled using the publicly accessible Laboro Tomato dataset, while the strawberry dataset was acquired from the publicly available StrawDI dataset. The technical path was divided into two objectives: model training, and model quantization and deployment. In the first, the RTFD model's performance was improved using the CIoU bounding box loss function, the ACON-C activation function, and the three-layer LC-PAN architecture.

The RTFD model was quantitatively trained for fruit detection. After being transformed into a Paddle Lite model and integrated into a testing Android smartphone app, the RTFD model performed extremely accurately in terms of real-time detection.

It is anticipated that edge computing will successfully implement the idea of redesigning the model structure, loss function, and activation function, as well as training by quantization, to expedite the detection of deep neural networks. The proposed RTFD has enormous potential for intelligent picking machines. For the strawberry and tomato datasets, PicoDet-S has an average accuracy of 94.2% and 76.8%, respectively. It is anticipated that edge computing will successfully implement the idea of redesigning the model structure, loss function, and activation function, as well as training by quantization, to expedite the detection using deep neural networks. The proposed RTFD has enormous potential for intelligent picking machines.

In line with Pereira et al. [34], six grape types that predominate in the Douro Region were automatically identified and classified using a methodology based on the AlexNet architecture and transfer learning scheme. Two natural vineyard image datasets, taken in various parts of Douro, were called Douro Red Grape Variety (DRGV), and GRGV\_2018. For picture managing, different image processing (IP) methods were applied, such as independent components filter (ICFs), leaf segmentation algorithm (LSA) with four-corners-in-one, leaf patch extraction (LPE), LPE with ICF, LPE with canny edge detector (CED), and LPE with Gray-scale morphology processing (GMP). These new datasets, with pre-processed and augmentation pictures were then trained in the AlexNet CNN.

The suggested method, four-corners-in-one, supplemented by the leaf segmentation algorithm (LSA), revealed success in reaching the best classification accuracy in the set of performed experiments. With a testing accuracy of 77.30%, the experimental results indicated the suggested classifier to be trustworthy. The algorithm took roughly 6.1 ms to identify the grape variety in a picture.

## Fruit Detection in Areas of Crops

Santos et al. [35] estimated grape wine production from RGB photos including deep learning algorithms and computer vision models. Pictures were taken of five distinct grape varieties, using a Canon camera and a smartphone. Mask R-CNN, YOLOv2, and YOLOv3 models from deep learning (DL) were trained to recognize and separate grapes in the photos. After that, spatial registration was carried out using the Structure from Motion (SfM) image processing technique, incorporating the information produced by the CNN-based stage. To prevent counting the same clusters across many photos, the clusters found in distinct images were removed using the CV model's outputs in the final phase.

While the Mask R-CNN outperformed YOLOv2 and YOLOv3 in terms of object detection, the YOLO model outperformed it in terms of detection time. Using YOLOv3, the poorest performance was attained. With an intersection over union (IoU) of 0.300, Mask R-CNN achieved an average accuracy of 0.805, a precision of 0.907, a recall of 0.873, and an F1-score of 0.890. YOLOv2 achieved an average accuracy of 0.675, a precision of 0.893, a recall of 0.728, and an F1-score of 0.802. In last place, YOLOv3 achieved an average accuracy of 0.566, a precision of 0.901, and a recall of 0.597, and an F1-score of 0.718.

In Assunção et al. [36], for a real-time peach fruit identification application, a tensor processing unit (TPU) accelerator was created with a Raspberry Pi target device, to give a lightweight and hardware aware MobileDet detector model. Three fruit peach cultivars— Royal Time, Sweet Dream, and Catherine—were combined into one picture dataset. A RGB camera was used to capture the pictures. The following components make up the hardware platform (edge device) utilized to execute inferences: a Raspberry Pi 4 microcontroller development kit; a Raspberry Pi Camera Module 2, a Coral TPU accelerator, a DC-to-DC converter, and three Li-ion batteries. As a detector, the single-shot detector (SSD) model was applied. The backbones underwent SSD modifications. In this paper, a MobileNet CNN was used as the basis for the SSD

model in experiments to look at the trade-off between detection accuracy and inference time. MobileNetV1, MobileNetV2, MobileNet EdgeTPU, and MobileDet were the backbones that were utilized.

In comparison to the other models, SSD MobileDet excelled, achieving an average precision of 88.2% on the target TPU device, according to the data. The model with the least performance degradation (drop) was SSD MobileNet Edge TPU, which had a decrease of 0.5%; the model with the most impact, SSD MobileNetV2, experienced a drop of 1.5%. SSD MobileNetV1 has the smallest latency at 47.6 ms (average). The authors have contributed to the field by expanding the applications of accelerators (the TPU) for edge devices in precision agriculture.

---

## References

1. Alibabaei, K.; Gaspar, P.D.; Lima, T.M.; Campos, R.M.; Girão, I.; Monteiro, J.; Lopes, C.M. A review of the challenges of using deep learning algorithms to support decision-making in agricultural activities. *Remote Sens.* 2022, 14, 638.
2. Afonso, M.; Blok, P.M.; Polder, G.; van der Wolf, J.M.; Kamp, J. Blackleg detection in potato plants using convolutional neural networks. *IFAC-Pap.* 2019, 52, 6–11.
3. Assuncao, E.; Diniz, C.; Gaspar, P.D.; Proenca, H. Decision-making support system for fruit diseases classification using Deep Learning. In *Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DAS A)*, Sakheer, Bahrain, 8–9 November 2020; pp. 652–656.
4. Azgomi, H.; Haredasht, F.R.; Mottagh, M.R.S. Diagnosis of some apple fruit diseases by using image processing and artificial neural network. *Food Control* 2023, 145, 109484.
5. Kerkech, M.; Hafiane, A.; Canals, R. Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Comput. Electron. Agric.* 2020, 174, 105446.
6. Sujaritha, M.; Annadurai, S.; Satheeshkumar, J.; Sharan, S.K.; Mahesh, L. Weed detecting robot in sugarcane fields using fuzzy real time classifier. *Comput. Electron. Agric.* 2017, 134, 160–171.
7. Milioto, A.; Lottes, P.; Stachniss, C. Real-time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. *arXiv* 2018, arXiv:1709.06764.
8. Lottes, P.; Behley, J.; Milioto, A.; Stachniss, C. Fully Convolutional Networks with Sequential Information for Robust Crop and Weed Detection in Precision Farming. *IEEE Robot. Autom. Lett.* 2018, 3, 2870–2877.
9. Ma, X.; Deng, X.; Qi, L.; Jiang, Y.; Li, H.; Wang, Y.; Xing, X. Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PLoS ONE* 2019, 14, e0215676.
10. Ferreira, A.D.S.; Freitas, D.M.; da Silva, G.G.; Pistori, H.; Folhes, M.T. Unsupervised deep learning and semi-automatic data labeling in weed discrimination. *Comput. Electron. Agric.* 2019, 165, 104963.
11. Wang, A.; Xu, Y.; Wei, X.; Cui, B. Semantic Segmentation of Crop and Weed using an Encoder-Decoder Network and Image Enhancement Method under Uncontrolled Outdoor Illumination. *IEEE Access* 2020, 8, 81724–81734.
12. Kamath, R.; Balachandra, M.; Vardhan, A.; Maheshwari, U. Classification of paddy crop and weeds using semantic segmentation. *Cogent Eng.* 2022, 9, 2018791.
13. Mu, Y.; Feng, R.; Ni, R.; Li, J.; Luo, T.; Liu, T.; Li, X.; Gong, H.; Guo, Y.; Sun, Y.; et al. A Faster R-CNN-Based Model for the Identification of Weed Seedling. *Agronomy* 2022, 12, 2867.
14. Assunção, E.; Gaspar, P.D.; Mesquita, R.; Simões, M.P.; Alibabaei, K.; Veiros, A.; Proença, H. Real-Time Weed Control Application Using a Jetson Nano Edge Device and a Spray Mechanism. *Remote Sens.* 2022, 14, 4217.
15. Peña, J.; Torres-Sánchez, J.; Serrano-Pérez, A.; de Castro, A.; López-Granados, F. Quantifying Efficacy and Limits of Unmanned Aerial Vehicle (UAV) Technology for Weed Seedling Detection as Affected by Sensor Resolution. *Sensors* 2015, 15, 5609–5626.
16. Huang, H.; Deng, J.; Lan, Y.; Yang, A.; Deng, X.; Zhang, L. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. *PLoS ONE* 2018, 13, e0196302.
17. Bah, M.; Hafiane, A.; Canals, R. Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. *Remote Sens.* 2018, 10, 1690.
18. Osorio, K.; Puerto, A.; Pedraza, C.; Jamaica, D.; Rodríguez, L. A Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images. *AgriEngineering* 2020, 2, 471–488.
19. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo Algorithm Developments. *Procedia Comput. Sci.* 2022, 199, 1066–1073.

20. BioD'Agro. E 3.3 Arquitetura, Desenvolvimento e Testagem do Algoritmo de Análise de Dados. BioD'Agro Project Report. March 2023. Available online: <https://biodagro.wearespaceway.com/biblioteca-e-eventos/entreg%C3%A1veis> (accessed on 13 April 2023). (In Portuguese).
21. Islam, N.; Rashid, M.M.; Wibowo, S.; Xu, C.Y.; Morshed, A.; Wasimi, S.A.; Moore, S.; Rahman, S.M. Early Weed Detection Using Image Processing and Machine Learning Techniques in an Australian Chilli Farm. *Agriculture* 2021, 11, 387.
22. Dyrmann, M.; Karstoft, H.; Midtiby, H.S. Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 2016, 151, 72–80.
23. Andrea, C.-C.; Daniel, B.B.M.; Misael, J.B.J. Precise weed and maize classification through convolutional neuronal networks. In Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), Salinas, Ecuador, 16–20 October 2017; pp. 1–6.
24. Gao, J.; Nuytens, D.; Lootens, P.; He, Y.; Pieters, J.G. Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. *Biosyst. Eng.* 2018, 170, 39–50.
25. Bakhshipour, A.; Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput. Electron. Agric.* 2018, 145, 153–160.
26. Sa, I.; Chen, Z.; Popović, M.; Khanna, R.; Liebisch, F.; Nieto, J.; Siegwart, R. weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming. *IEEE Robot. Autom. Lett.* 2018, 3, 588–595.
27. Yang, X.; Ye, Y.; Li, X.; Lau, R.Y.K.; Zhang, X.; Huang, X. Hyperspectral image classification with deep learning models. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 5408–5423.
28. Yashwanth, M.; Chandra, M.L.; Pallavi, K.; Showkat, D.; Kumar, P.S. Agriculture Automation using Deep Learning Methods Implemented using Keras. In Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 6–8 November 2020; pp. 1–6.
29. Jin, X.; Che, J.; Chen, Y. Weed Identification Using Deep Learning and Image Processing in Vegetable Plantation. *IEEE Access* 2021, 9, 10940–10950.
30. El-Kenawy, E.S.M.; Khodadadi, N.; Mirjalili, S.; Makarovskikh, T.; Abotaleb, M.; Karim, F.K.; Alkahtani, H.K.; Abdelhamid, A.A.; Eid, M.M.; Horiuchi, T.; et al. Metaheuristic Optimization for Improving Weed Detection in Wheat Images Captured by Drones. *Mathematics* 2022, 10, 4421.
31. Sunil, G.C.; Koparan, C.; Ahmed, M.R.; Zhang, Y.; Howatt, K.; Sun, X. A study on deep learning algorithm performance on weed and crop species identification under different image background. *Artif. Intell. Agric.* 2022, 6, 242–256.
32. Sunil, G.C.; Zhang, Y.; Koparan, C.; Ahmed, M.R.; Howatt, K.; Sun, X. Weed and crop species classification using computer vision and deep learning technologies in greenhouse conditions. *J. Agric. Food Res.* 2022, 9, 100325.
33. Mao, D.; Sun, H.; Li, X.; Yu, X.; Wu, J.; Zhang, Q. Real-time fruit detection using deep neural networks on CPU (RTFD): An edge AI application. *Comput. Electron. Agric.* 2022, 204, 107517.
34. Pereira, C.S.; Morais, R.; Reis, M.J.C.S. Deep Learning Techniques for Grape Plant Species Identification in Natural Images. *Sensors* 2019, 19, 4850.
35. Santos, T.T.; de Souza, L.L.; dos Santos, A.A.; Avila, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Comput. Electron. Agric.* 2020, 170, 105247.
36. Assunção, E.; Gaspar, P.D.; Alibabaei, K.; Simões, M.P.; Proença, H.; Soares, V.N.; Caldeira, J.M. Real-Time Image Detection for Edge Devices: A Peach Fruit Detection Application. *Future Internet* 2022, 14, 323.