MDPI

Review

Methods for Detecting and Classifying Weeds, Diseases and Fruits Using AI to Improve the Sustainability of Agricultural Crops: A Review

Ana Corceiro [1], Khadijeh Alibabaei [2], Eduardo Assunção [1,3,4], Pedro D. Gaspar [1,3,4,*] and Nuno Pereira [4]

[1] Department of Electromechanical Engineering, University of Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal; ana.corceiro@ubi.pt (A.C.); eduardo.assuncao@ubi.pt (E.A.)
[2] Steinbuch Centre for Computing, Zirkel 2, D-76131 Karlsruhe, Germany; khadijeh.alibabaei@kit.edu
[3] C-MAST Center for Mechanical and Aerospace Science and Technologies, University of Beira Interior, 6201-001 Covilhã, Portugal
[4] Department of Computer Science, Instituto de Telecomunicações, University of Beira Interior, 6201-001 Covilhã, Portugal; nuno.pereira@ubi.pt
[*] Correspondence: dinis@ubi.pt

Abstract: 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. However, agriculture's varied environments pose a challenge to testing and adopting new technology effectively. This study reviews recent advances in deep learning models and methods for detecting and classifying weeds to improve the sustainability of agricultural crops. The study compares performance metrics such as recall, accuracy, F1-Score, and precision, and highlights the adoption of novel techniques, such as attention mechanisms, single-stage detection models, and new lightweight models, which can enhance the model's performance. The use of deep learning methods in weed detection and classification has shown great potential in improving crop yields and reducing adverse environmental impacts of agriculture. The reduction in herbicide use can prevent pollution of water, food, land, and the ecosystem and avoid the resistance of weeds to chemicals. This can help mitigate and adapt to climate change by minimizing agriculture's environmental impact and improving the sustainability of the agricultural sector. In addition to discussing recent advances, this study also highlights the challenges faced in adopting new technology in agriculture and proposes novel techniques to enhance the performance of deep learning models. The study provides valuable insights into the latest advances and challenges in process systems engineering and technology for agricultural activities.

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

1. Introduction

The world population began growing rapidly during the industrial revolution, mainly due to medical advances and increases in agricultural productivity. Currently, it is estimated that the population increases dramatically, by an average of 80 million per year. With any type of projection, there is a degree of uncertainty; however, the United Nations predicts that there will be 8.1 billion people on the planet by 2025 [1–3].

In this era, the highest rates of population growth occur mainly in developing countries reflecting their higher fertility rates and an increase in longevity. Furthermore, urbanization

has been increasing, and by 2050 over 70% of the world's population will live in urban areas, dependent on food produced by others [2].

Given the increasing demand for food, the level of production should increase by 70% [1,3]. However, agriculture faces tremendous challenges including climate change, drought, pests, weeds, diseases, pollution, soil deterioration, pump irrigation costs, rising groundwater, the switch from a fuel-based to a bio-based transition economy, and finally, the decreasing availability of freshwater as demand rises [4].

According to [4], weeds directly compete with fruit or vegetable crops for water, growing space, nutrients, and sunlight, leaving the crops susceptible to insects and diseases, which results in productivity losses of 34% on average.

Water is one of the most important resources in crop growth. Amongst other issues, weeds can steal water from the field and prevent crop growth. Furthermore, climate change is directly connected with the changes in average global temperature. This contributes to the reduction of available water and, therefore, producers need to adopt water-saving practices, and keep the fields free of weeds [5].

One of the most essential factors in agricultural yield is weed management, and hand weeding is the oldest method. However, it has a high labor cost, and is inefficient and time-consuming. Mechanical weeding techniques are far more effective and labor-saving than hand weeding; however, they can easily cause crop damage [4].

One of the solutions for weed control is the application of herbicides. Although these chemicals can eliminate weeds efficiently, they can also pollute water supplies or food. Thus, to neutralize these problems, many European nations have started to restrict the use of pesticides in farming [6].

Due to these problems, weed management must become more environmentally friendly. Precision agricultural technology needs to be used to minimize the negative effects of herbicides on the environment and to optimize their usage [7,8].

Precision agriculture (PA) is the most favorable key to these problems. Using a variety of cutting-edge information, communication, and data analysis approaches, PA is a management strategy that aids in improving crop output while minimizing water and fertilizer losses, as well as enabling a better environmental impact [9].

The development of decision-making algorithms has placed a lot of emphasis on artificial intelligence (AI). AI includes any method that allows robots to learn from experience, adjust to new inputs, and emulate human behavior [10]. The sub-field of artificial intelligence known as machine learning (ML) employs computational algorithms to transform data from the real world into usable models and decision-making guidance. Finally, deep learning (DL) is a section of machine learning [11].

Remote sensing has been used frequently to map weed patches in agricultural fields for Site-Specific Weed Management (SSWM). Weeds can be identified or separated from cultivated plants based on their distinctive spectral signatures. Over the past few years, picture categorization using machine learning methods has proven to be very accurate and effective for weed mapping [9].

Convolution Neural Networks (CNNs) are now the most widely used deep learning method for the agriculture industry. Convolutional neural networks belong to a class of deep neural networks, usually employed to analyze visual imagery [12].

It is expected that these technologies will change agribusiness since they allow for decision-making in days rather than weeks. Additionally, they guarantee a significant drop in expenses and an increase in productivity [13].

In conclusion, to identify new approaches, difficulties, and potential fixes for employing deep learning in agriculture, this study will evaluate published studies. The objective of this paper is to give an overview of current work and to point out difficulties in the collection and preparation of data for deep learning models; to review the current DL model methods used in agriculture; to emphasize the challenges in model training; to evaluate the novel edge devices utilized to implement the trained models, as well as the

difficulties involved in using them in the real world. This review will not cover vegetation recognition methods based on radar/microwave sensors.

The remainder of this paper is structured as follows. Materials and methods, including information about the eligibility of the articles reviewed, are presented in Section 2. Section 3 gives an overview of deep learning as well as a review of the papers. The major challenges in this research are detailed in Section 4, while Section 5 concludes the article.

2. Materials and Methods

Since the evolution of weeds and the uncontrolled use of herbicides are hot and rising topics of the present times, a comprehensive literature review was conducted to collect, verify, analyze, and describe the scientific facts on the goals, difficulties, and constraints of weed detection and categorization, while putting people at the center of productive processes and systems.

As a result, the purpose of this systematic review is to research the philosophies of and approaches to deep learning in agriculture for purposes of detection and classification, with a focus directed to the problem of weeds.

Finally, this study was carried out using a four-phase flow diagram and PRISMA's standards, also known as a systematic review and meta-analysis statement.

2.1. Focus Questions

Deep learning emerges as a complement to agricultural production as a resource centered on the human being, where fruit and vegetable production is prioritized, to maintain productive and sustainable performance in respect of the supply of healthy foods [10].

This technique, due to its benefits, including strong feature extraction ability and excellent identification accuracy, is commonly employed in image recognition.

Future prospects for food generation are to reduce the number of plants that emerge by themselves in regions where the population maintains their crops. However, the objective is to do this sustainably and efficiently, contributing to the reduction of the uncontrolled use of herbicides.

That said, it is important to research and evaluate the deep learning approaches, introducing and evaluating the ideas and philosophies behind weeds, their detection, and classification, to achieve a sustainable and resilient system, especially for the worker. This led us to our research questions:

(1)　What types of models can be applied for deep learning for detection or classification in agriculture?
(2)　Which model is better for detection and/or classification?
(3)　What type of metrics can be used to evaluate the model?

2.2. Sources of Information and Methods Used to Obtain Data

Initial data screening and collection for this systematic literature review article began in September 2022, and for the bibliographic study, three electronic databases, Science Direct, ResearchGate, and Google Scholar, were employed. Pre-determined keywords connected to the study's main emphasis were utilized for the database search: weed detection and weed classification. The keywords were chosen to be comprehensive and not to condition or restrict the study.

Therefore, all information and data that would be pertinent to the inquiry were included. The screening titles' keywords included "deep learning" together with one of the following keywords: "agriculture", "weeds", "precision agriculture", "weeds detection", "weeds classification", and "crop classification". The language search was always conducted in English.

2.3. Eligibility Criteria

In this analysis, the evolution of distinct strategies for weed, disease, and fruit detection and classification are explored, as well as their classification employing performance metrics. Based on article titles and abstracts, the writers initially performed a preliminary selection and exclusion process. For eligibility, the following inclusion and exclusion criteria were used. Only research published in English with text available, including research articles and review articles, was included. Furthermore, given the recent development of the technologies for weed detection and classification under consideration here, a study period between 2015 and 2022 was chosen.

In addition to these inclusion factors, articles should describe and explore at least one of the two focus subjects of the study: artificial intelligence and weeds detection/classification. Moreover, other features were added such as results, sustainability, and the algorithms associated with deep learning for detection or classification. Additionally, some of the other applications include the detection of plant and leaf diseases and detection and classification of fruits or vegetables were introduced. Articles before 2015, with a non-exclusive focus on classification or detection, and those not considering deep learning were excluded. The following aspects were considered while analyzing the articles:

- The process used to gather the dataset and the difficulties encountered when using it to train the model.
- The performance of the models and the DL models/architectures employed in the paper.
- The measures that were used for the model's evaluation.
- The model's inference time (if specified), as this is a crucial factor in the use of the model in real-time applications.
- The examination of the model's failure prediction.
- Whether the trained model was deployed using a low-cost device developed by the authors.

3. Review of Extracted Research

3.1. Principal Findings

To perform this review, a four-phase flow diagram PRISMA were applied. Thus, the procedure employs four phases, of which the first is the identification of the papers. The second phase is the screening. Following this, the eligibility is considered, and the number of papers is determined.

This literature research obtained an aggregate of 175 articles: 35 from Science Direct, 53 from ResearchGate, and 87 from Google Scholar. Of these articles, 12 were not available for full-text reading and 59 were duplicates or triplicates among the three databases, and for that reason were excluded.

Thus, when the non-eligible articles were excluded, 104 articles remained. Upon reviewing the titles and abstracts of the papers, 62 papers were eliminated in the following stage, leaving 42 articles. The 42 publications were then subjected to a full-text analysis to determine their eligibility, during which 9 were disqualified since they did not match the current study's objectives (Figure 1).

This literature study comprised analysis of the remaining 33 publications.

Dates and numbers of papers were the focus of the first examination. As can be seen in Figure 2, despite the publications' recent publication dates (from 2015 to 2022), there has been a marked increase in research into the notion of deep learning for weed categorization and detection.

The second study goal was to validate the genre of recently themed articles that were being written and distributed. Among the 33 articles, four dealt with plant diseases, four with fruit detection, fourteen with weed detection, and eleven with weed categorization. This is represented in Figure 3.
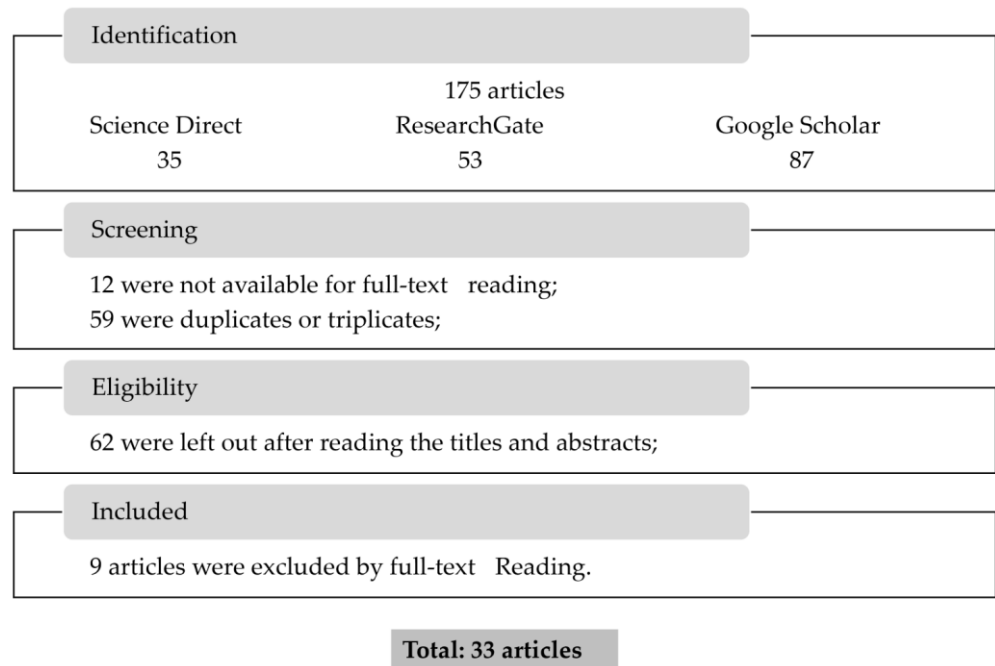
Figure 1. Diagram created using PRISMA that shows the systematic research's findings.
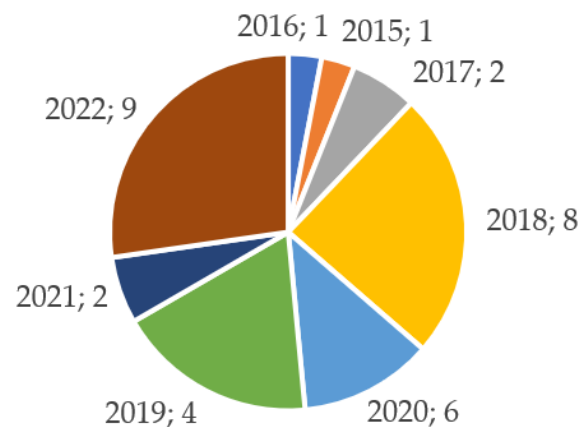

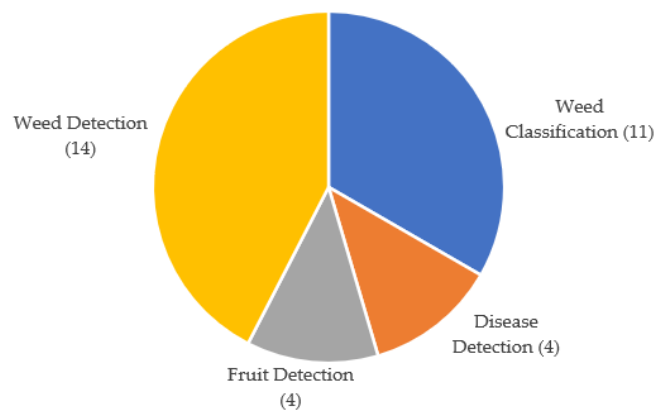**Figure 2.** Number of articles published by year (Year; Number of publications).

Figure 3. Papers published by application area.

3.2. An Overview of Deep Learning

The brain's neural networks in humans served as an inspiration for DL models. The phrase "deep" refers to how many covert levels the data is converted through. To produce

predictions, the models pass the input across a deep network with several layers, each of which employs a hierarchy to extract certain characteristics from the data at various sizes or resolutions and combine them into a higher-level feature [11]. DL models are broken down into three categories: reinforcement learning, unsupervised learning, and supervised learning. The process of learning a function using labelled training data is known as supervised learning. In supervised learning, each pair of data in the dataset represents the intended output value and an input item. The purpose of reinforcement learning (RL) is to determine how an agent should behave in each environment to maximize rewards.

A model of supervised learning is the CNN model, created primarily for segmentation, classification, and detection.

Convolutional layers, pooling layers, nonlinear operations, and fully linked layers make up a CNN model. A convolutional layer is a method for obtaining characteristics from inputs and has a series of kernels, the parameters of which must be learned. The dot product between each kernel entry and each point in the input is calculated as each kernel is moved over the input's height and width [11].

A nonlinear activation function is utilized to include nonlinear characteristics in the model after each convolutional layer. Rectified Linear Units (ReLU) are the most popular activation function in DL models since they are state-of-the-art. Between two convolutional layers, a pooling layer is typically employed to decrease the number of parameters and prevent overfitting [11]. The subsequent layers are made up of completely interconnected layers that use the features that the preceding layer retrieved to provide class probabilities or scores. All the neurons in these layers are linked to this layer [11].

3.2.1. Segmentation by CNN

CNN models can also be used to complete the segmentation task. Segmentation is the division of an image into groups of pixels and the assignment of a class to each group. DeepLab, Mask R-CNN, and Fully Convolutional Networks (FCN) are DL models for segmentation [11].

3.2.2. Detection by CNN

The CNN model may also be used for object detection. You Only Look Once (YOLO), Single Shot Multi-box Detector (SSD), and LedNet are examples of single-stage detectors. Current object detection algorithms come in two varieties. Two-stage detectors include R-CNN, Fast R-CNN, and Faster R-CNN, for instance. The first stage of the two-stage detector establishes regions of interest using the Region Proposal Network (RPN). The suggested regions are then put to the test for item classification using bounding box regression and convolutional layers. In contrast, bounding boxes and object categorization are instantly provided by a single feed-forward convolutional network in single-stage detectors [11].

3.2.3. Classification by CNN

Regression or classification problems can be resolved using the CNN model. When performing classification or regression tasks, the last layer of the model is chosen as a completely connected layer with a SoftMax activation function, and as a fully connected layer, frequently with a linear function. AlexNet, GoogleNet, VGG, and ResNet are among the most often used CNN designs for classification, along with more contemporary, lighter models such as MobileNet and EfficientNet [11].

3.2.4. Vegetation Index

In precision agricultural applications involving remote sensing, vegetation indices (VIs) are frequently utilized. In both qualitative and quantitative vegetation analysis, they are regarded as being particularly useful for tracking the development and health of crops. Vegetation indices are based on the vegetation's ability to absorb electromagnetic radiation [13]. There are modifications in the mathematics of electromagnetic spectrum scattering and absorption in various bands depending on the type of vegetation. and

these parameters can be determined using data from either each individual shot or after the creation of orthophotos showing the whole crop. Several factors, including plant biochemical and physical characteristics, ambient factors, soil background characteristics, and moisture content impact the reflectance in different bands. Differences in reflectance can provide accurate temporal and geographical information on the crops being monitored [13].

There are two primary categories for vegetation indices: vegetation indices created on hyperspectral or multispectral data and those based on information from the visible spectrum [13]. The capacity to identify bands of radiation of green (G), healthy vegetation has been greatly enhanced by the development of simple vegetation indicators that can integrate RGB (Red-Green-Blue) data with various spectral bands, such as NIR (Near-infrared), and RE (Red Edge) [13].

The Ratio Vegetation Index (RVI), and Normalized Difference Vegetation Index (NDVI), which are based on the NIR and bands of radiation of Red (R) channels of radiation, are intended to give more pronounced contrast between the plant and the soil. The RVI, shown in Equation (1), is one of the multispectral plant indices, which highlights the difference between soil and vegetation. It is also susceptible to the visual characteristics of the ground. The NDVI, shown in Equation (2), is a development of RVI and is determined by the visible and near-infrared light reflected from the vegetation. It is the most well-known and often used vegetation index. It provides a simple way to monitor the development and health of many agricultural crops, since unhealthy or sparse vegetation reflects more visible light and less near-infrared radiation [13].

The ratio of NIR to RE radiation is normalized by the Normalized Difference Red Edge (NDRE), shown in Equation (3). The Green Normalized Difference Vegetation Index (GNDVI) with NIR and Green (G) bands, is shown in Equation (4).

The most used indices when analyzing the VIs generated from RGB photos are Excess Greenness Index (ExG) and Normalized Difference Index (NDI). ExG is predicated on the idea that plants exhibit a superior level of greenness, and that soil is the only background component. It is determined by the doubling the radiation in G channels minus the radiation in Red (R) and Blue radiation (B) channels, as shown in Equation (5). The NDI was suggested to use just green and red channels to differentiate plants from background pictures of dirt and debris, as shown in Equation (6). Table 1 provides a selection of the most used vegetation indices [13].

Table 1. Most applied vegetation indices.

Vegetation Index	Abbreviation	Formula	N°
Vegetation Indices derived from multispectral information			
Ratio Vegetation Index	RVI	$\frac{NIR}{R}$	(1)
Normalized Difference Vegetation Index	NDVI	$\frac{NIR-R}{R+NIR}$	(2)
Normalized Difference Red Edge Index	NDRE	$\frac{NIR-RE}{RE+NIR}$	(3)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{NIR-G}{NIR+G}$	(4)
RGB-based Vegetation Indices			
Excess Greenness Index	ExG	$2 \cdot G - R - B$	(5)
Normalized Difference Index	NDI	$\frac{G-R}{G+R}$	(6)

3.2.5. Data Acquisition

A large amount of labelled data is necessary for DL-based weed identification and classification approaches. In the first step, it is important to acquire a useful quantity of data for further analysis. Image acquisition can be defined as the act of obtaining an image from different sources. Hardware systems such as cameras, encoders, and sensors can be used for this. Different images can be captured depending on the type of camera, and the type of sensor embedded. The sensors' job is to take pictures with great temporal

and spatial resolution, which can help with identifying a variety of vegetation-related characteristics [14]. Different types of sensors deployed on a range of platforms are used to capture various modalities of data. These pictures can be captured through unmanned aerial vehicles (UAV), field robots (FR), all-terrain vehicles (ATV), micro aerial vehicles (MAV), cameras, satellite images, and public datasets [7].

Datasets can also be obtained from free internet sources, such as the Broadleaf Dataset, DeepWeed [15], Oilseed image, and Sugar Beets [16]. Each type of sensor can monitor many aspects of the plant, including its color, texture, and geometric shape. Some sensors can measure specific wavelengths of radiation. The information gathered by these sensors may be further processed to track critical agricultural properties during the various growth stages, including soil moisture, plant biomass, and vegetation health [13].

Nowadays, for agriculture, there are four types of sensors entrenched in cameras: hyperspectral, multispectral, RGB, and thermal sensors. Visible light sensors are the most popular sensors for PA applications. They are economical, easy to use, and can capture high resolution photographs. In addition, the information obtained needs straightforward processing [13].

RGB sensors are frequently adjusted to obtain information on radiation in other bands, most commonly the Infrared (IR) or RE band. This is created by changing one of the original optical filters for one that allows the perception of NIR [13].

In addition to collecting data using wavelengths that are visible, multispectral devices also gather data using wavelengths that are invisible to the human eye, such as near-infrared radiation, short-wave infrared radiation (SWIR), and others [3].

Hyperspectral sensors analyze a wide spectrum of light instead of just assigning primary colors to each pixel, using wavelengths with a range of 400 to 1100 nm in steps of 1 nm. Both multispectral and hyperspectral sensors can obtain data regarding the vegetation's spectral absorption and reflection on numerous bands [13].

Thermal infrared sensors detect the temperature of the materials and produce pictures using this information instead of their visible properties. Thermal cameras employ infrared sensors and an optical lens to collect infrared energy. In these cases, warmer items are frequently shown as yellow and cooler ones as blue. However, this type of sensor is used for very specific applications, for example, irrigation management [13]. Other types of sensors, described in the literature, can be employed in addition to the ones indicated above, such as light detection and range (LiDAR) sensors. These devices are also known as "distance sensors", and when used in conjunction with other sensors they could be used by vehicles (such as robots) to navigate in the field [13].

3.2.6. Performance Metrics

The quality or efficiency of the model is assessed using measurement techniques referred to as performance metrics or evaluation metrics. With the help of these performance indicators, how well the simulation processed the given data can be assessed. It is feasible to improve the performance of the model by changing the hyper-parameters [17]. The most used metrics to determine the excellence of the model are shown in Table 2.

Table 2. Principal performance metrics used to evaluate the models.

Performance Metric	Formula	N°
Precision	$\frac{TP}{TP+FP}$	(7)
Recall	$\frac{TP}{TP+FN}$	(8)
True Negative Rate	$\frac{TN}{TN+FP}$	(9)
F1-Score	$2\frac{P \cdot R}{P+R}$	(10)
Kappa Coefficient	$\frac{P_a - P_r}{1 - P_r}$	(11)
Normalized mutual information	$\frac{I(X,Y)}{\sqrt{H(X) \cdot H(Y)}}$	(12)

The confusion matrix is a tabular representation of the ground-truth labels and model predictions [18]. In the matrix, columns correspond to the predicted values, and rows specify the actual values. Both ground truth and predicted values have two possible classes, positive or negative. An assessment factor is represented by each cell in the confusion matrix: True Positive (TP) denotes the number of positive class samples that the model correctly predicted; True Negative (TN) is the number of negative class samples that the model properly predicted; False Positive (FP) is the number of negative class samples that the model erroneously predicted, while False Negative (FN) denotes the number of positive class samples that the model incorrectly predicted [7,19].

Accuracy (ACC) corresponds to a percentage of correctly predicted events divided by all predicted events; this means that it is the degree of closeness to the true value [7].

Precision (P) is the degree to which an instrument or process will repeat the same value. It is the number of TP values, from all relevant findings, divided by the total number of TP and false positives (FP), as shown in Equation (7) [7,17].

Recall (R), Positive Rate (TPR) or sensitivity, is fundamentally the percent of real positives compared to all the ground truth's positives. It is the number of TP values, divided by the sum of TP and FN values, as shown in Equation (8) [17,20].

Specificity or True Negative Rate (TNR) is a test's capacity accurate to identify negative results. It is the number of TN values, divided by the sum of TN and FP values, as shown in Equation (9) [19,20].

The F1-score is the harmonic mean between recall and precision. It is the fraction given by the multiple of P and R divided by the sum of P and R, multiplied by two, as shown in Equation (10) [19,20].

The Kappa Coefficient (k) measures the degree to which the projected values and the real values accord [7]. Cohen's Kappa is calculated as the probability of agreement (Pa) minus the chance of random agreement (Pr) and then divided by one minus the probability of random agreement, as shown in Equation (11), [7].

The area under the receiver operating characteristic curve (AU-ROC or AUC) represents a graph displaying a classification model's effectiveness at different threshold levels. The sensitivity and false-positive rate (FPR) are shown on a probability curve called the ROC. The AUC calculates the performance across the thresholds and provides an aggregate measure [21].

Intersection over union (IoU) is a metric used to estimate how well a predicted mask or bounding boxes match the ground truth data, by dividing the area of overlap by the area they cover as a union [17].

Mean Intersection over Union (mIoU) is the dataset's average IoU for each class of an item [7].

Normalized mutual information (NMI) is the accepted metric for assessing clustering outcomes [15]. This statistic may be used to trade off the number of clusters vs the quality of the clustering. Using two random variables (X and Y), the NMI is determined by Equation (12). In this equation, H is entropy, and I is the mutual information metric. This measurement is done by contrasting the labels assigned to the clusters with actual labels [15]. Table 2 provides a selection of the most used performance metrics [7,11].

3.3. Brief Review of Papers

3.3.1. 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 [11].

Disease Detection in Individual Fruits

Afonso et al. [22] 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 our binary classification utilizing logarithmic SoftMax activation (healthy versus blackleg).

Assunção et al. [23] 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 our 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. [24], 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. Figure 4 presents an input of an apple fruit, with both healthy and infected parts. After processing, the affected area is shown as orange in the middle picture, and the healthy area is then painted in yellow and the affected area in black, in the right-hand photo.

Disease Detection in Areas of Crops

Table 3 describes the features of research works in the field of disease detection.

In Kerkech et al. [25], 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.
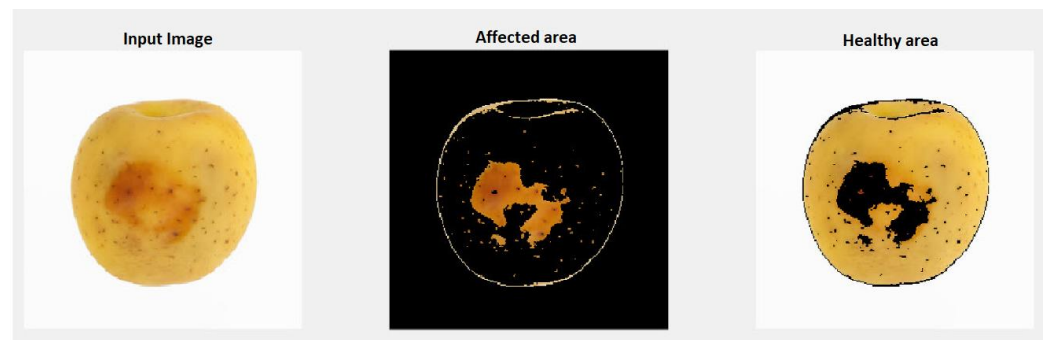

Figure 4. Example of the input fruit, showing the separated infected and healthy parts [24].

Figure 5 shows an example of segmentation by SegNet, and the fusion compared with the ground truth (GT). The first set of images (a–h) does not show examples of the symptom class, so it is healthy; it can be shown that the visible and infrared estimates and the fusion are similar. However, in the second set (i–p), it can be observed that the ground truth in both spectra, which depicts a region that is almost entirely polluted by mildew, is the same except for the distinct color code.
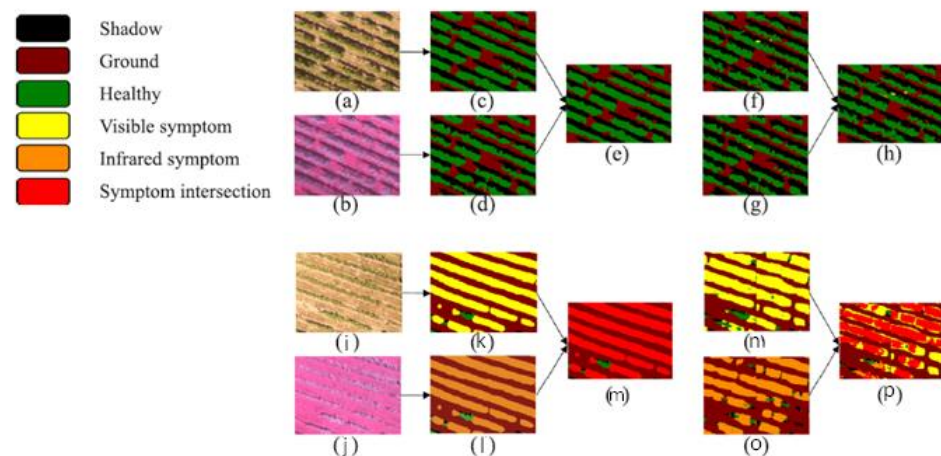

Figure 5. An illustration of segmenting and fusing a healthy region. Images in the following order: (**a**) visible image, (**b**) infrared image, (**c**) visible GT, (**d**) infrared GT, (**e**) fusion GT, (**f**) visible SegNet estimate, (**g**) infrared SegNet estimation, and (**h**) fusion of segmentation findings. An example of segmenting and fusing a mold-infested region. For example, (**i**) stands for visible image, (**j**) for infrared image, (**k**) for visible ground truth, (**l**) for infrared ground truth, (**m**) for fusion ground truth, (**n**) for visible segmentation net estimation, (**o**) for infrared segmentation net estimation, and (**p**) fusion of segmentation results (adapted from [25]).

Table 3. Feature descriptions of publications in the field of "Disease Detection".

References	Application	Data Used	Model Used	Metric Used	Model Performance
Afonso et al. [22]	Classify diseased potato plants.	RGB camera.	Deep CNN: ResNet18, ResNet50.	Precision Recall	The findings of this study demonstrate that a CNN, more especially ResNet18, may function as a reliable detector for potatoes infected with the blackleg disease in the field. The detection performance can also be anticipated to increase with larger datasets and data augmentation.
Assuncão et al. [23]	Detect peach disease.	Six datasets from [23].	CNN	F1-score	No disease class is incorrectly classified by the model. These successes highlight the promise of CNN for classifying fruit diseases with little training data. The model is also made to work with portable electronics.
Azgomi et al. [24]	Detect apple disease.	Digital camera.	MLP ANN	Accuracy	The results showed a maximum accuracy of 73.7% for the implementation of a two-layer structure with eight neurons in the first layer and eight neurons in the second layer.
Kerkech et al. [25]	Detect Esca disease in grapevine.	UAV system with an RGB sensor.	CNN: SegNet.	Accuracy	In comparison to the model trained using infrared pictures, the RGB pictures gave better performance. One of the research's flaws is the small size of the training sample, which negatively impacted how well the deep learning segmentation worked.

3.3.2. 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 [11].

Weed Detection in Individual Plants

In accord with Sujaritha et al. [26], 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. [27] 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. [28] 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. [29] 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.

The proposed method worked effectively in classifying the pixels in pictures of weeds and shaped rice seedlings found in paddy areas. The U-Net and FCN techniques had an

average accuracy rate of 89.5% and 70.8%, respectively. Figure 6 shows the experimental results for the FCN based on SegNet and U-Net compared with the original and ground truth images; blue represents rice, brown represents weeds, and the grey scale is the background.
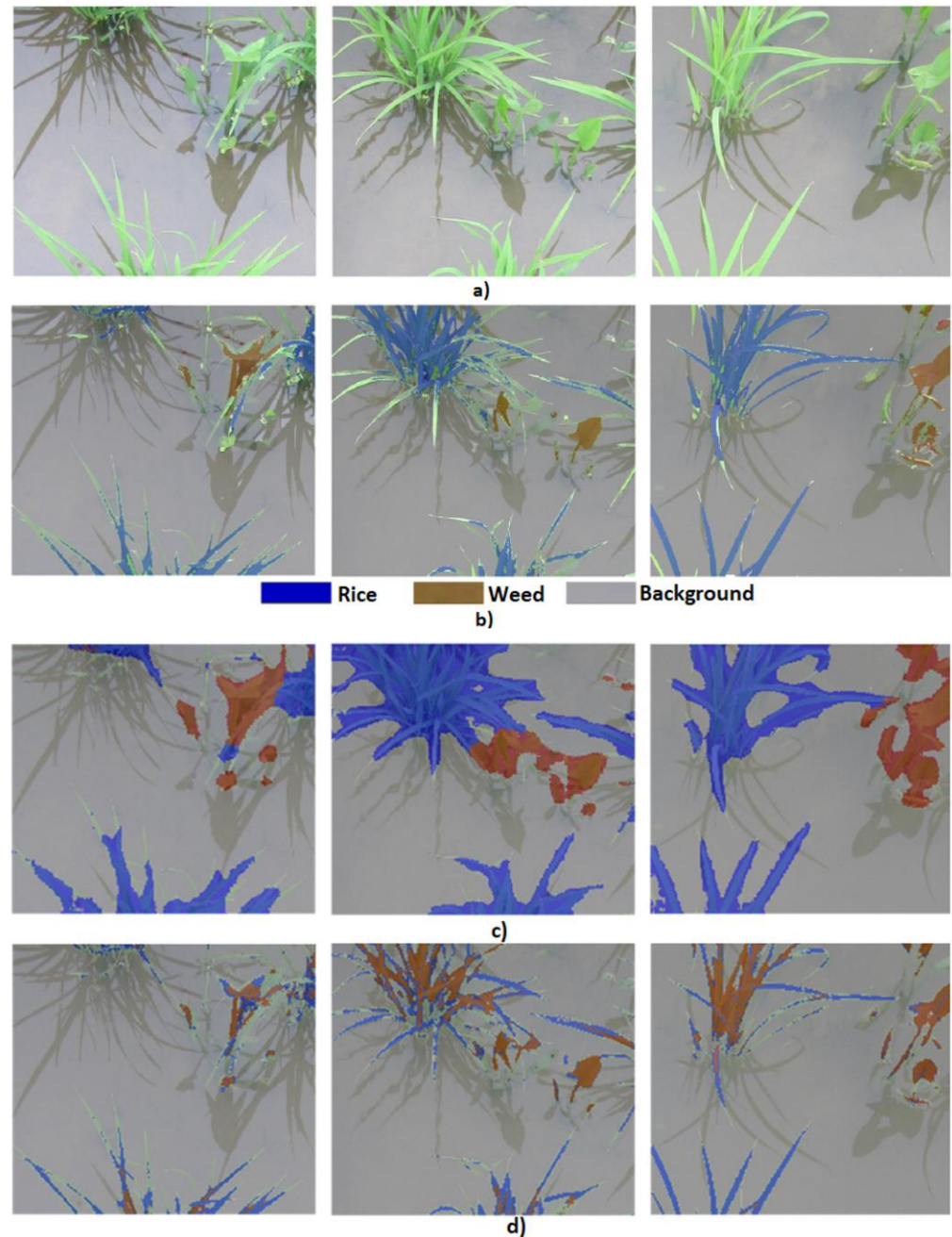

Figure 6. Results of the proposed model. (**a**) Original images; (**b**) Ground truth images; (**c**) Output by FCN; (**d**) Output by U-Net [29].

Ferreira et al. [15] 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. [16], 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. [30] 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 upscaled 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. Figure 7 represents the results of the proposed model using PSPNet; the images of the first row correspond to the original images, the second row represents the predicted output, and the last one is the ground truth image. The first line relates to the paddy, and the second characterizes the broadleaved weed. After that, the third shows the broadleaved weed (blue) and paddy (yellow), and the fourth, the sedge weeds. Sedge weeds were difficult to identify, whereas broadleaved weeds and paddy were clearly identified. This loss could be explained by how sedges and paddy are alike.
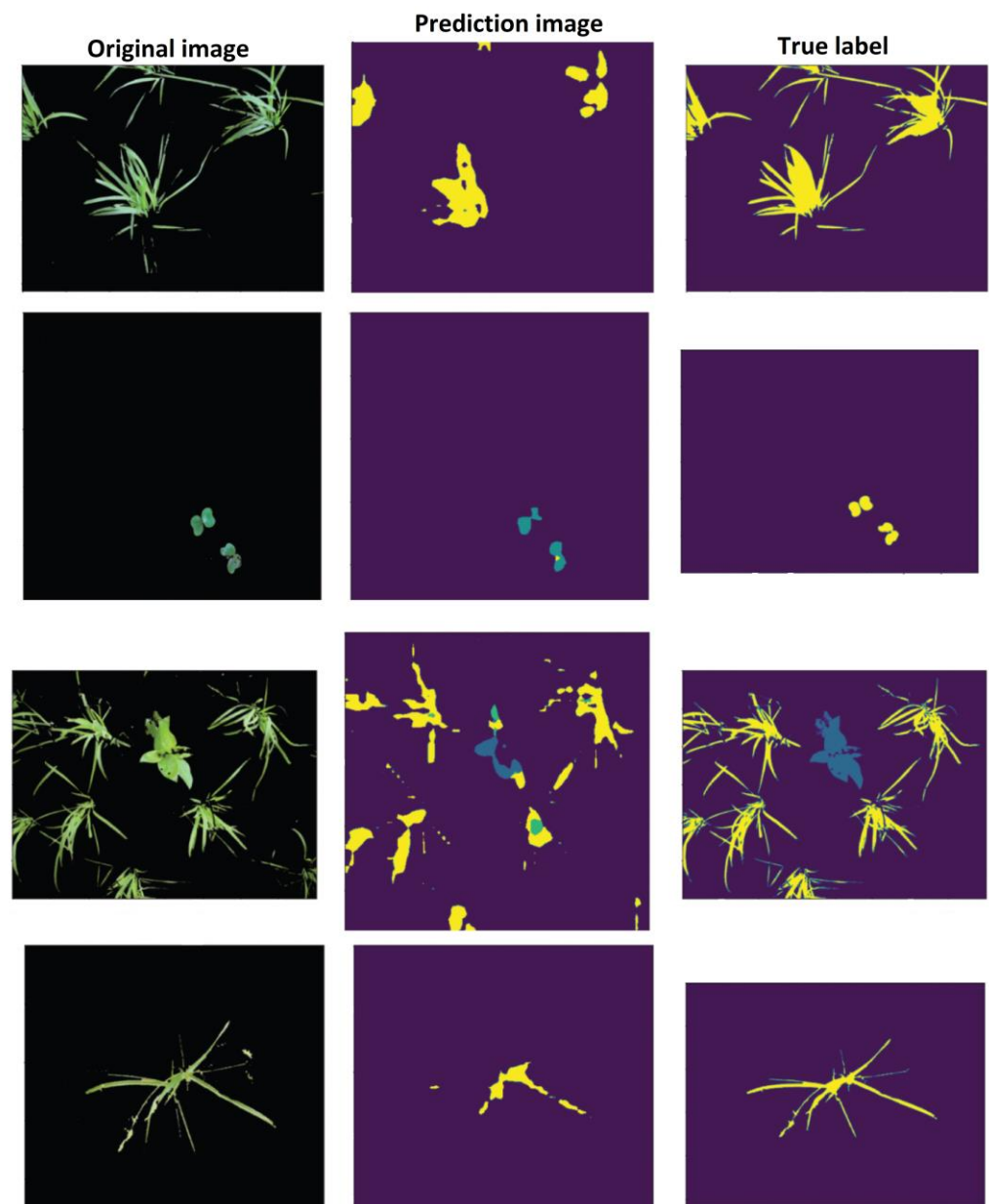
Figure 7. The results of PSPNet, the images of the first row, correspond to the original images, while the second row represents the predicted output, and the last is the ground truth image. The first line is the paddy, the second one is the broadleaved weed, the third one is the broadleaved weed (blue) and paddy (yellow), and the last, is the sedge weed, (adapted from [30]).

Mu et al. [31] 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. [32] 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 MobilenetV2 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.

In addition, with the CWFID and weeds dataset, the outcomes were also contrasted with the initial segmentation work. The test with OS = 8 and DM = 1.0 achieved a mIOU of 75%, and an OS = 32 and DM = 0.5 accomplished a mIOU of 64%. Figure 8 displays the relevant segmentation quality outcomes. Different hyperparameters for DM and OS caused variations in segmentation performance (quality).

In Figure 9, a version of non-weeds segmentation is in the upper-left corner. The subsequent pictures display the input weed picture along with the associated segmentation outcomes (output). In the middle of each segmented region are the green dots that represent the weeds' center of gravity. The findings demonstrate the method's viability and outstanding spraying precision. Given that the trade-off between segmentation accuracy and inference time can be managed via the hyperparameters DM and OS, DeepLabV3 has shown to be an incredibly flexible model for segmentation tasks. Weed spraying in real-time was also precise and practical. The system correctly positioned the nozzle at all target weeds and sprayed the spray, as seen in the video demonstration. This outcome demonstrates the possibility of improvements for creating compact models with high predictive accuracy.
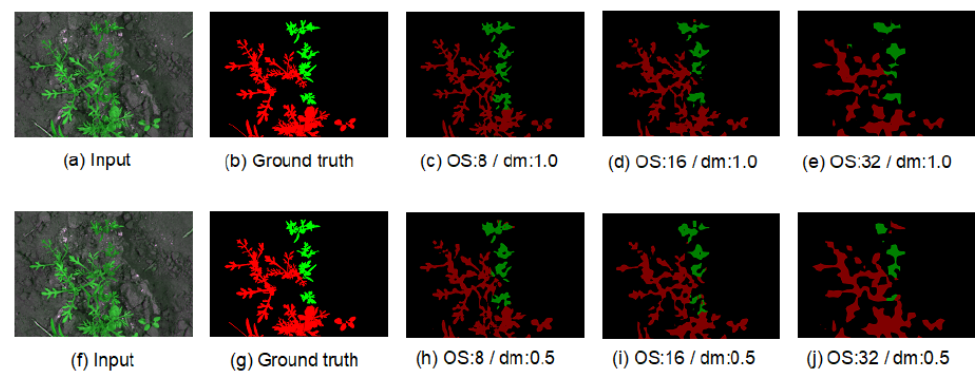
Figure 8. CWFID and weeds dataset qualitative segmentation results. Except for input and ground truth, the labels of each sub-image indicate the optimization that was applied to the model to produce the segmentation [32]. (**a**) Input; (**b**) Ground truth; (**c**) OS:8/dm:1.0; (**d**) OS:16/dm:1.0; (**e**) OS:32/dm:1.0; (**f**) Input; (**g**) Ground truth; (**h**) OS:8/dm:0.5; (**i**) OS:16/dm:0.5; (**j**) OS:32/dm:0.5.
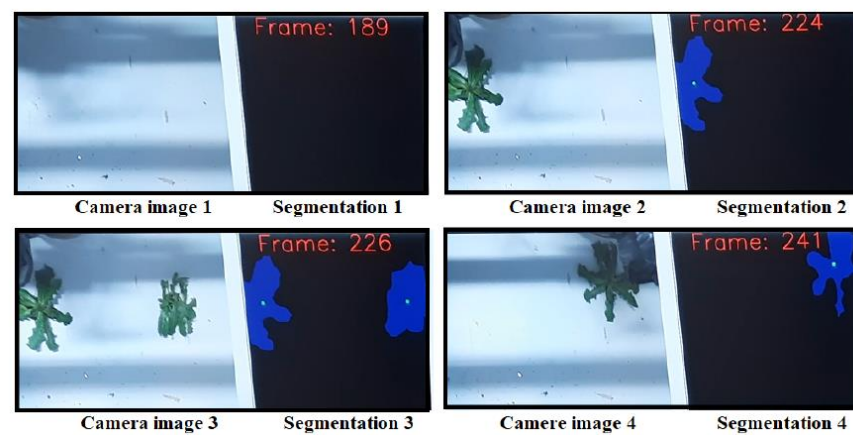

Figure 9. Segmentation of weeds as result of the application in real time [32].

Weed Detection in Areas of Crops

Peña et al. [33] 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. Figure 10 shows a comparison of results. The first line (A) presents on-ground photographs, while the second line (B) shows manual categorization of observed data. The third line

(C) shows the image classification achieved by the OBIA algorithm. The model results were divided into four types: the number of correct frames (1); underestimated weeds (2), namely, frames with weed infestations in which the OBIA system spotted some weed plants but missed others; false negative frames (3), with weed-infested frames in which no weeds were detected; and false positive frames (4), in which weeds were overestimated.
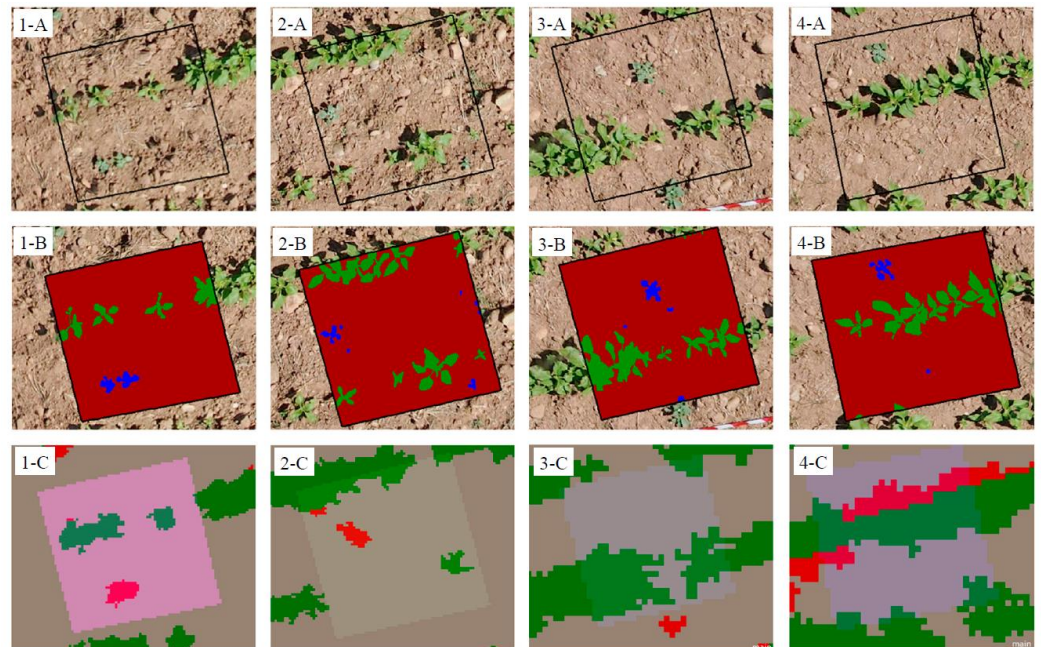

Figure 10. Example of the four sample frames' outcomes. (**A**) On-ground images; (**B**) manual classification of observed data; and (**C**) image classification conducted by the OBIA algorithm. (**1**) Correct categorization; (**2**) underestimation of weeds; (**3**) negative errors; (**4**) false positive errors [33].

Huang et al. [34] 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. In Figure 11, it is possible to observe the results of the FCN with different pre-trained CNNs. In (a), the real UAV images are presented; (b) is the ground truth representation; and images from (c) to (e) show results obtained by FCN-AlexNet, FCN-VGG16, and FCN-GoogLeNet, respectively.

Bah et al. [35] 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.
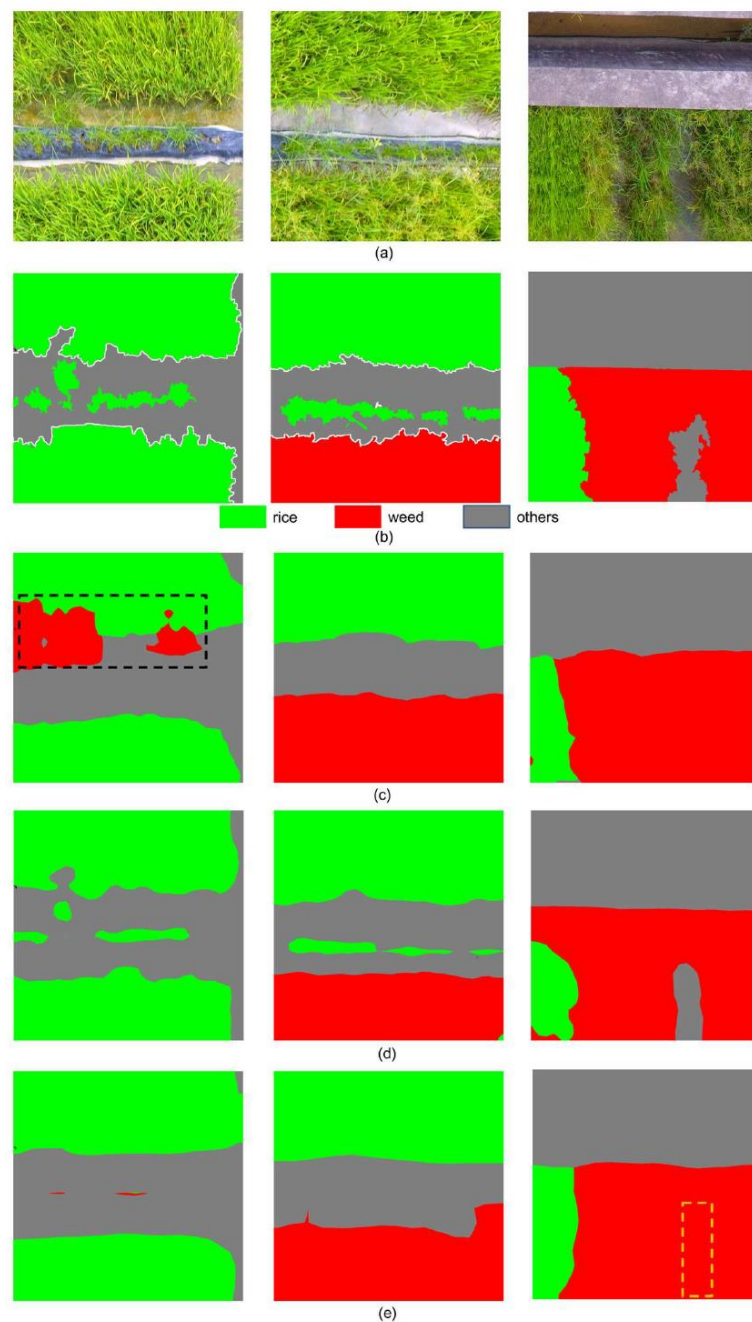
Figure 11. Classification results of the FCN with distinct pre-trained CNNs. (**a**) Real UAV image; (**b**) Ground truth results; (**c**–**e**) Results acquired by FCN-AlexNet, FCN-VGG16 and FCN-GoogLeNet, respectively [34].

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. Figure 12 shows two examples of image classification with models produced by unsupervised data in spinach fields at the top (a,b), and bean fields at the bottom (c,d). The samples acquired using a sliding window, without a crop line or any background details, are shown on the left (a,c). The weeds found after applying crop line and background information are shown on the right in red (b,d). The plants are designated as being crops, weeds, or ambiguous decisions by the red, blue, and white dots, respectively.
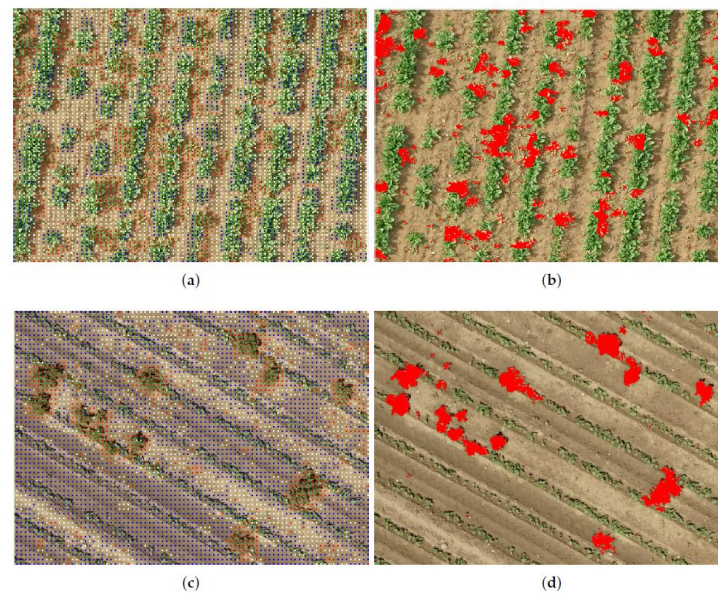
Figure 12. Examples of unmanned aerial vehicle (UAV) picture categorization using unsupervised data models in spinach, and bean fields. (**a**) Sample of a spinach field acquired using a sliding window, without a crop line or any background details; (**b**) Sample from spinach field acquired after applying crop line and background information; (**c**) Sample of a been field acquired using a sliding window, without a crop line or any background details; (**d**) Sample from been field acquired after applying crop line and background information. Crops are represented in blue, weeds in red, and uncertain decisions in white [35].

In line with Osorio et al. [8], 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 [36].

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 [36].

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 [37].

Islam et al. [14] 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.

Table 4 describes the feature of research works in the field of Weed Detection.

3.3.3. 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 [11].

Weed Classification in Individual Plants

According to Dyrmann et al. [38], 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 Midtiby, Søgaard (2005), Minervini, Scharr, Fischbach, and Tsaftaris (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.

Table 4. Feature descriptions of publications in the field of "Weed Detection".

References	Application	Data Used	Model Used	Metric Used	Model Performance
Sujaritha et al. [26]	Weeds in sugar cane fields' detection.	Two digital cameras.	Fuzzy real time classifier.	Accuracy	The prototype distinguished between nine different weed species with greater precision and a shorter processing time. The field obstructions provoked the robot to deviate from the planned path while it is moving.
Milioto et al. [27]	Crop-weed sugar beet detection (VIs added to the input).	4-channel RGB and NIR camera.	Mask R-CNN.	Accuracy IoU Precision Recall	By including a new channel at its input, the model's performance improved. Compared to the NIR channel network, the RGB channel network converges 15% more quickly to the final accuracy of 95%.
Lottes et al. [28]	Crop and weed detection in the sugar field.	FR with a 4-channel RGB+NIR camera.	Encoder–Decoder FCN: DenseNet.	F1-score	The suggested model was able to robustly identify crops in all growth phases and outperformed encoder-decoder FCN without a sequential model, RF, and vanilla FCN with an F1-score of 92.3.
Ma et al. [29]	Detection of weeds and rice seedlings.	RGB images.	FCN: SegNet.	Accuracy	The suggested technique successfully classified the pixels in images of weeds and distinctively shaped rice seedlings discovered in paddy regions. The SegNet approach produced a good accuracy classification.
Ferreira et al. [15]	Detection of weeds in a soybean field (unsupervised clustering).	Two datasets from [15].	JULE; DeepCluster.	Accuracy NMI	Performance-wise, DeepCluster outperformed JULE. The outcomes from these datasets point to a viable use of clustering and unsupervised learning for agricultural issues.
Wang et al. [16]	Detection of crops of sugar, weeds, and oilseeds.	Two datasets from [16].	Encoder-decoder CNN.	Accuracy IoU	The outcomes show how useful NIR information is for exact segmentation in low-light settings. The accuracy of segmentation was greatly increased by the addition of NIR data.
Kamath et al. [30]	Weed detection in paddy crops.	Digital camera.	PSPNet, UNet, and SegNet.	IoU	In terms of efficiency, PSPNet fared better than SegNet and UNet. The frequency weighted IoU falls between 80% and 90%, with the mean IoU lying between 70% and 80%.

Table 4. *Cont.*

References	Application	Data Used	Model Used	Metric Used	Model Performance
Mu et al. [31]	Weed identification in maize, sugar beet, and wheat crops.	Dataset: V2 Plant Seedlings, from [31].	FPN; Faster R-CNN: ResNeXt.	Accuracy Recall F1-Score IoU	The experimental findings demonstrate that by merging the ResNeXt feature extraction network with the FPN network, the Faster R-CNN-FPN deep network model achieves a higher recognition accuracy.
Assunção et al. [32]	Weed detection using semantic segmentation.	Five datasets from [32].	TMG DeepLabV3 MobilenetV2.	mIOU	Given that the trade-off between segmentation accuracy and inference time can be managed via the hyperparameters OS and DM, DeepLabV3 has been shown to be an incredibly flexible model for segmentation tasks.
Peña et al. [33]	Weed seedlings in sunflower field detection.	Visible-light and multispectral cameras in a UAV.	OBIA.	Accuracy	While the visible light camera performed better at lower flight altitudes, the multispectral camera proved more accurate at higher altitudes. The spectrum mixing of flowers and bare soil components caused some mistakes in the higher elevations.
Huang et al. [34]	Weed cover maps to detect weeds and crops in rice fields.	UAV with a digital camera.	FCN.	Accuracy IoU	The FCN technology performed well in terms of efficiency and accuracy for weed identification. Since FCN is a supervised algorithm, it necessitates a lot of manual labelling effort because it needs a large amount of labelled pictures for training and updating.
Bah et al. [35]	Weed detection in bean and spinach fields.	Drone with a digital camera.	CNN: Resnet18.	Accuracy	Given the disparities between supervised and unsupervised labelling's accuracy, the unsupervised one may be a preferable option for weed detection, especially when crop rows are widely spaced.
Osorio et al. [8]	Weed detection in lettuce crops (VIs added to the input).	Mavic Pro with a multispectral camera.	SVM+ HOG; Mask R-CNN; YOLOv3.	F1-Score Accuracy Precision Recall Specificity	The HOG-SVM approach was shown to work quite well, and given that it requires less processing power, it is an excellent choice for IoT systems. As compared to the other two, the YOLO approach overestimates the high values of weed coverage.
Islam et al. [14]	Weed detection.	RGB camera coupled in a UAV.	RF; KNN; SVM.	Accuracy Recall Precision FPR Kappa	The experimental results indicate that RF outperformed the other classifiers. The efficiency of RF and SVM as classifiers for weed detection from UAV pictures is noteworthy.

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. [39] 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. [40] 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 [41], 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. The results of the SBWD algorithm at various levels are shown in Figure 13. The initial RGB image is shown in (a); (b) shows the EXG method for segmenting plants; (c) demonstrates the image created using

morphological techniques (noise removal, area thresholding for removing small plants, and edge erosion for removing touching overlaps); (d) shows the SBWD algorithm for segmenting sugar beets; (e) shows the subtraction of the result of greenness from image (c) from image (b) showing the weeds; and, finally (f) displays the result of the SBWD algorithm showing weeds, sugar beet, and false negatives; red pixels indicate weeds, green pixels indicate sugar beet plants, and yellow pixels show areas that were incorrectly identified as undesirable objects.
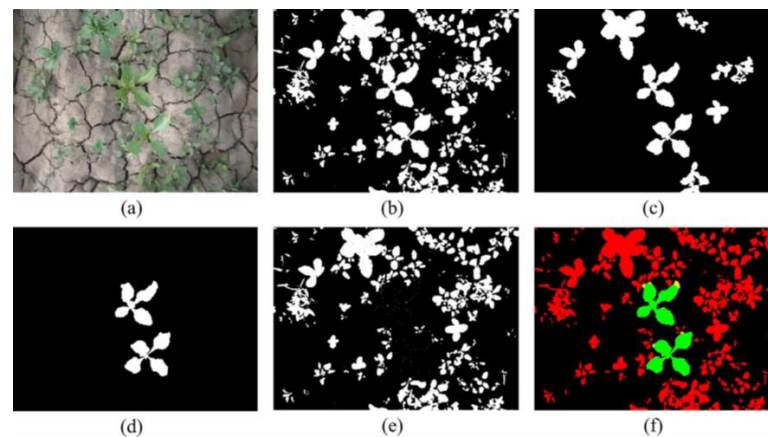

Figure 13. Results from steps of the SBWD algorithm: (**a**) initial RGB image; (**b**) segmented plants using EXG method; (**c**) the result after morphological filtering of small objects; (**d**) segmented sugar beets with SBWD algorithm; (**e**) subtraction of image (**c**) from image (**b**) showing the weeds; (**f**) result of SBWD algorithm showing weeds, sugar beet and false negatives [41].

Sa et al. [42] performed weed and sugar beet classification using a CNN with multi-spectral 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. [43] 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. [44] 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. [45] 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. [46], 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. [47] 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. [48] 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%.

Table 5 summarizes the features of research works in the field of weed classification.

3.3.4. 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. [49] 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.

Table 5. Feature descriptions of publications in the field of "Weed classification".

References	Application	Data Used	Model Used	Metric Used	Model Performance
Dyrmann et al. [38]	Weed classification in 22 different crops.	Six databases from [38].	CNN.	Accuracy	In general, the classes with the highest number of species also had the greatest categorization accuracy. As a result, classes with fewer picture samples made a smaller total loss. The network's classification precision varied from 33% to 98%, with an average precision of 86.2%.
Andrea et al. [39]	Image segmentation and classification of weeds in maize fields.	Multispectral camera to acquire RGB and NIR images.	CNN: LeNet, AlexNet, cNet, and sNet.	Accuracy	Based on its accuracy and processing speed, the network cNET provided the greatest training outcomes.
Gao et al. [40]	Weed and maize classification (VIs added to the input).	Hyperspectral snapshot camera sensor.	RF.	Precision Recall F1-score	The RF model that was used to create classifiers using various spectral feature combinations performed well. Vegetation indices are useful techniques for developing important aspects for the categorization of crops and weeds.
Bakhshipour and Jafari. [41]	Classification of sugar beet crop and weeds.	RGB camera.	SVM; ANN.	Accuracy	Both ANN and SVM properly identified the effectiveness of sugar plants and weeds.
Sa et al. [42]	Classification of sugar beet and weeds.	Multispectral images collected by a MAV.	CNN.	Accuracy F1-score	Most of the weeds were classified well. Due to restrictions in the dataset the model was trained on, certain spatiotemporal discrepancies were identified.
Yang et al. [43]	Classification of weeds in crops and landscapes.	Six datasets from [13].	CNN: 2-D-CNN, 3-D-CNN, R-2-D-CNN, R-3-D-CNN.	Accuracy	For most of the data sets, the suggested R-3-D-CNN model performs better than most of the current models and can also converge more quickly. Nevertheless, compared to conventional machine learning techniques, these models need more training samples.
Yashwanth et al. [44]	Image classification of weeds in nine different crops.	Digital camera.	Keras API; TensorFlow.	Accuracy	The model was tested using nine different types of crops and the corresponding weeds, and the greatest accuracy was determined to be 96.3%. All the provided photos were correctly categorized as either plants or weeds.

Table 5. *Cont.*

References	Application	Data Used	Model Used	Metric Used	Model Performance
Jin et al. [45]	Weed identification in cabbage fields.	Digital camera.	CNN: CenterNet.	Precision Recall F1-score	The recommended approach has application value for the sustainable development of the vegetable sector and is suited for ground-based weed detection in vegetable agricultural land under diverse circumstances, lighting, and complicated backdrops, as well as various growth phases.
El-Kenawy et al. [46]	Weed classification in wheat crops.	Images captured by a drone.	NN; SVM; KNN; GWO; SCA.	Accuracy F1-score Recall Specificity	The research shown that the suggested strategy outperforms existing methods and improves 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%.
G C et al. [47]	Weed classification in sugar beets.	Four Canon digital cameras.	CNN: VGG16, ResNet50.	F1-Score	The VGG16 and ResNet50 models, which were created using non-uniform backdrop photos, performed well on the uniform background, with average f1-scores of 82.75% and 75%, respectively. Employing non-uniform backgrounds led to poorer results. The model that was trained using combined datasets from two background scenarios performed better than any.
Zhang et al. [48]	Weed classification in black bean, canola, corn, flax, soybean, and sugar beets.	RGB camera.	SVM; VGG16.	F1-score	All SVM model classifiers have failed in comparison to the VGG16 model classifiers. The results demonstrated that the range of the VGG16 model classifier's average F1-scores was between 93% and 97.5%. The range of SVM average F1 scores was 83 to 94 percent. In the VGG16 Weeds-Corn classifier, the corn class scored 100% F1.

Figure 14 shows the results of strawberry and tomato detection. The picture contains varied colored borders that reflect separate categories. The blue arrows serve as suggestive indication symbols, and the blue circles highlight regions of faulty or missing detections. The red, orange/yellow, and light blue correspond, respectively, to mature strawberries, half-mature strawberries, and immature strawberries.

Figure 14. The impact of RTFD model detection on the tomato and strawberry datasets. The blue circles denote locations of inaccurate or missing detections, while the blue arrows act as suggestive indicator symbols. The red, orange/yellow, and light blue corresponded, respectively to mature strawberries, half-mature strawberries, and immature strawberries [49].

In line with Pereira et al. [50], 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. [51] estimated grape wine production from RGB photos including deep learning algorithms and computer vision models. Pictures were taken of five distinct grape varietals, 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.

Figure 15 shows an example of the detection of the five grape varieties with the three neural networks employed, viz., Mask R-CNN, YOLOv2, and YOLOv3, as well as the ground truth images. In the image it is possible to observe several object identification results, where the color does not indicate correlation. In this example, it is possible to observe the difference between the models and better understand visually the results from performance metrics.

Figure 15. Example of object detection results produced by the three neural networks, namely, Mask R-CNN, YOLOv2, and YOLOv3, compared to the ground truth images, for the five grape varieties in the study. The same color does not mean correspondence [51].

In Assunção et al. [52], 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. Figure 16 shows an example of detection samples of the three cultivars, with Catherine at the left, Sweet Dream in the middle, and Royal Time at the right.

Table 6 summarizes the features of research work in the field of fruit detection.

Table 6. Feature descriptions of publications in the field of "Fruit Detection".

References	Application	Data Used	Model Used	Metric Used	Model Performance
Mao et al. [49]	Identify tomatoes and strawberries.	Two datasets: StrawDI; Laboro Tomato, from [21].	PicoDet-S.	Accuracy	The proposed RTFD has enormous potential for intelligent picking machines, and 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.
Pereira et al. [50]	Identify and classify grapes.	Two datasets: DRGV And DRGV_2018, from [23].	LSA; CED; GMP; LPE; ICF; and CNN: AlexNet.	Accuracy	With a testing accuracy of 77.30%, the experimental findings proved the suggested classifier's trustworthiness. The algorithm took roughly 6.1 ms to identify the grape variety in a picture.
Santos et al. [51]	Identify grapes and estimate grape wine yield.	RGB camera.	Mask R-CNN; YOLOv2; YOLOv3.	Precision Recall F1-score	The YOLO model beat the Mask R-CNN in terms of detection time, while the Mask R-CNN outperformed YOLOv2 and YOLOv3 in terms of object detection. Using YOLOv3, the poorest performance was attained.
Assunção et al. [52]	Identify peaches.	RGB camera.	MobileDet; MobileNet Edge TPU; MobileNetV2; MobileNetV1.	Precision	The model performed at 19.84 frames per second (FPS) with an average precision (AP) of 88.2% and a 640×480 picture size. According to the results, the TPU accelerator can be a great replacement for processing at the cutting edge in precision agriculture.

Figure 16. Detection sample for the Catherine peach cultivar (**left**), Sweet Dream peach cultivar (**middle**), and Royal Time peach cultivar (**right**), (adapted from [52]).

4. Discussion

According to the research analyzed here, the following conclusions of each study as well as specific challenges can be indicated.

In ref. [8], the authors employed three models, namely, SVM-HOG, Mask R-CNN, and YOLOv3. Images were captured by a multispectral camera and used the following performance metrics: F1-Score, accuracy, specificity, precision, and recall. The HOG-SVM approach was shown to work quite well, and given that it requires less processing power, it is an excellent choice for IoT systems.

In ref. [14], the authors used three models, namely, RF, KNN, and SVM. Images were captured by a RGB camera and used the following performance metrics: accuracy, k, FPR, precision, and recall. The findings of this study indicate that RF outperformed the other classifiers. Furthermore, the efficiency of RF and SVM as classifiers for weed detection from UAV pictures is noteworthy.

Another model was employed in ref. [15], with JULE and DeepCluster. In this study, images were taken from two datasets, and the authors choose accuracy and normalized mutual information. The model achieved better performance with DeepCluster. Furthermore, the outcomes from these datasets point to a viable use of clustering and unsupervised learning for agricultural issues.

In ref. [16], an encoder-decoder CNN was employed, with pictures from two datasets, and working with accuracy and IoU to evaluate the model. The results demonstrate the effectiveness of NIR information for precise segmentation under low lighting conditions, while VIs without NIR information did not improve the segmentation results.

In ref. [22], the authors employed a deep CNN, with pictures from a RGB camera, and used the following performance metrics: precision and recall. The findings of this study demonstrate that a CNN, more especially ResNet18, may function as a reliable detector for potatoes infected with the blackleg disease in the field. However, with larger datasets and data augmentation, the performance can be increased.

In ref. [23], a CNN was used, with pictures from five datasets, and an F1-Score was applied as a performance metric. The findings of this study demonstrate that the CNN did not classify any disease incorrectly.

A MLP with an ANN model was developed in ref. [24], pictures were taken with a digital camera, and the model was evaluated in terms of accuracy. The implementation of a two-layer structure with eight neurons in the first layer and eight neurons in the second layer produced a maximum accuracy of 73.7%.

In ref. [25], a CNN was employed, with accuracy as a performance metric. Images were captured by a RGB sensor. The findings of this study demonstrate that the model trained with RGB photos performed better than the model trained with infrared images. The limited size of the training sample is one of the research study's weaknesses.

A fuzzy real time classifier was created in ref. [26], and two digital cameras were utilized, while accuracy was used for classifying the model. The prototype distinguished

weed species with greater precision and a shorter processing time. However, the field obstructions provoked the robot to diverge from the planned path while it was moving.

A Mask R-CNN was applied in ref. [27], with RGB and NIR images; the model was evaluated through accuracy, IoU, precision, and recall. The model achieved better performance by adding an extra channel at the input of the model.

In ref. [28], an encoder- decoder CNN was evaluated, and the authors used RGB and NIR images, with the F1-Score as a classification metric. The suggested model was able to robustly identify crops in all growth phases and outperformed encoder-decoder FCN.

An FCN with RGB images was developed in ref. [29]. Once again, accuracy was used as a metric. The suggested technique successfully classified the pixels in images of weeds.

Four different models of the CNN were utilized in ref. [30], a digital camera was used, and this model was evaluated by IoU. The results showed that in terms of efficiency, PSPNet fared better than SegNet and UNet.

In ref. [31], two models were studied, FPN, and Faster R-CNN, where images were from a dataset, and accuracy, recall, F1-Score, and IoU were used as metrics. The experimental results show that the Faster R-CNN-FPN deep network model obtains greater recognition accuracy by employing the ResNeXt feature extraction network and combining the FPN network.

In ref. [32], five datasets were used in the following models: TMG, DeepLabv3, and MobileNetv2. The results show that the trade-off between segmentation accuracy and inference time can be managed via the hyperparameters OS and DM. DeepLabV3 has shown itself to be an incredibly flexible model for segmentation tasks.

In ref. [33], an OBIA model was used, with images from a visible-light and multi-spectral camera. Accuracy was used to evaluate the system. The findings of this study demonstrate that the multispectral camera was more accurate at higher flight altitudes, whereas the visible light one was better at lower altitudes. However, the spectrum mixing of flowers and bare soil components caused some mistakes at higher elevations.

In ref. [34], the authors employed an FCN, with pictures from a digital camera, and used the performance metrics of accuracy and IoU. The findings of this study demonstrate that the FCN technology performed well in terms of accuracy and efficiency for weed identification. On the other hand, it necessitates a great deal of manual labelling effort because it needs a large number of labelled pictures for training and updating.

In ref. [35], a CNN was developed, images were taken from a digital camera, and the authors used accuracy as a metric. In terms of adaptability and flexibility, this method is attractive since a model may be simply trained on a dataset. On the other hand, it necessitates a great deal of manual labelling.

In ref. [38], the authors used six datasets in a CNN, the model was evaluated using accuracy. The model achieved an accuracy of 86.2%. However, classes with fewer picture samples made a smaller total loss.

Different models of CNN were employed in ref. [39]. Images were taken by RGB and NIR cameras, and, once again, accuracy was used as the key metric. Based on its accuracy and processing speed, the network cNET provided the greatest training outcomes.

An RF model was developed in ref. [40]; the authors used hyperspectral images, and evaluated the model using precision, recall, and F1-Score. The results showed that the RF model performed well, and the vegetation indices are also useful techniques for developing important aspects of the categorization of weeds and crops.

In ref. [41], a SVM and an ANN were created with RGB images, using accuracy as a performance metric. The results showed that both models properly identified weeds and sugar plants.

In ref. [42], a CNN was designed, multispectral images were used, and the model was evaluated using accuracy, and F1-Score. The results showed that most of the weeds were classified well. However, spatiotemporal inconsistencies were found in the model due to limitations in the training dataset.

Four models of CNN were employed in ref. [43], and six datasets were used, as well as accuracy as a performance metric. The results showed that the proposed R-3-D-CNN model frequently outperforms existing models for most of the data sets and can also converge more quickly. However, these models require more training samples.

In ref. [44], a model with Keras API, and Tensorflow was implemented, and a digital camera was used to take the pictures. Accuracy was the evaluation metric. The crops and weeds were correctly detected, achieving the greatest accuracy.

In ref. [45], genetic algorithms were implemented with CenterNet. A digital camera was used as well as, and precision, recall, and F1-Score were applied to evaluate the model. The suggested method is suitable for ground-based weed identification in vegetable agricultural land under various conditions, illumination, complex backgrounds, as well as various growth stages.

In ref. [46], three models were developed, NN, SVM, and KNN, and these versions were evaluated with the genetic algorithms, GWO, and SCA. The research showed that the proposed technique performs better than other methods and increases classification accuracy.

A CNN model was implemented with CGG16, and ResNet50 in ref. [47]. Four digital cameras were used to take the images, and the model was evaluated by F1-score. The results showed that the model that was trained using combined datasets from two background scenarios performed best. Furthermore, the models which were built using non-uniform backdrop pictures behaved well on the uniform background. Those trained on a uniform background functioned poorly

In ref. [48], the researchers created a model with SVM, and VGG16, where RGB images were applied. The results showed that the VGG16 model classifiers outperformed all SVM model classifiers.

A Mask R-CNN, with a YOLOv2, and YOLOv3 were employed in ref. [51]. In this model, RGB images were used, and authors utilized precision, recall, and F1-Score to evaluate the model. The results showed that the YOLO model beat the Mask R-CNN in terms of detection time, while the Mask R-CNN outperformed YOLOv2 and YOLOv3 in terms of object detection.

A picoDet-S CNN model was engaged in ref. [49]. Two datasets were used, and accuracy was the performance metric. The proposed RTFD has enormous potential for intelligent picking machines, and 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.

In ref. [50], a CNN model with optimizing algorithms was developed. Two datasets were used, as the accuracy as a performance metric. The results demonstrate that the model classifiers are trustworthy with an accuracy of 77.30%.

In ref. [52], a MobileNet with TPU model was employed, the research used RGB images, and the model was evaluated using precision as a performance metric. According to the results, the TPU accelerator can be a great replacement for processing at the cutting edge in precision agriculture.

The papers under examination highlight the following difficulties. The most critical point involves the datasets. Even with transfer learning and data augmentation, training the model may still need a substantial quantity of data, and an insufficient dataset for training the model might result in substantial failures [22,25,43]. Datasets with more samples will be able to perform better. Furthermore, the quality of the datasets is also a problem; images with bad quality will perform worst [42]. As a result, the first and most crucial phase is gathering real field data and photos under various circumstances.

Agricultural image datasets are also more complex due to outdoor conditions, the fact that the object of interest typically occupies a very small and off-center portion of the image, the similarity between objects and background, the obstruction of the object by leaves and branches, the presence of multiple objects in one image, and a variety of other factors. The dataset must, however, accurately reflect the condition of the environment for it to be useful in the actual world [49,51,52]. Furthermore, datasets with LED pictures will

have a low accuracy [38]. Data augmentation may also be useful in some circumstances, such as fluctuating illuminance.

Another important challenge is the large amount of data that needs to be labelled. This task is expensive and time-consuming. Moreover, some tasks can only be carried out by experts in the industry, such as tasks involving plant diseases. Supervised learning needs a huge number of labelled images, for training and updating the models [34,35].

Data augmentation and transfer learning, as observed in multiple works, are approaches to avoid labelling a huge dataset, although labelling a small dataset still takes time. Unsupervised and semi-supervised learning techniques can be very beneficial but still require further research [12].

The performance of the model is impacted by the type of input [23,24,27,28,33]. The model's performance is affected by background removal from images [46], using various color spaces and vegetation indices as input [27], and crop detection at various growth stages [45]. The altitude when images are taken is also important for the input [33]. Finding the ideal input set for a given activity is therefore difficult.

In addition, field obstructions are a problem for the use of field robots. When robots find some obstacle in their path, they diverge from the planned path [26]. To solve this problem, instead of robots, drones can be applied. Another solution is that farmers can have the field clean and plain.

Accuracy and inference time must be traded off when selecting a model for a task. The model can be selected based on the application. In the field of agriculture, no setting is exactly comparable to another, and each environment and problem has its unique dataset; therefore, the DL model could not be relevant in all situations. The model performance could suffer because of the variations in the visual quality of the photos in the training and test datasets [38]. Retraining the already learned model using a tiny dataset from the new environment is one technique to get around the problem [11].

Moreover, the performance of these models depends on the choice of hyperparameters, loss functions, and optimization algorithms. Algorithms such as Bayesian optimization can help to find the right hyperparameters [11,23].

The models' capacity to be applied in real-time presents another difficulty. Most deep learning models need to be trained on many parameters, and once trained, the model's inference is not made in real time. Time inference is crucial in some applications, such as employing a robot for harvesting. However, there are still several issues with implementation on devices such as smartphones that must be considered, including memory usage and performance. Deep Learning models may now be used in practical and real-world applications because of the emergence of edge devices such as the Raspberry Pi and Jetson Nano, lightweight categorization models such as MobileNet, and cloud computing. The model size may be compressed, and the detection speed increased using the quantization approach [11,23].

5. Conclusions

The manuscript discusses the use of deep learning in agriculture and biodiversity and identifies certain difficulties in the field. It is suggested that reduced herbicide administration, minimal pesticide use, organic farming, suitable crop rotations, small-scale fields, and preservation of natural gaps between agroecosystems may contribute to more sustainable agriculture and the development of biodiversity in agricultural systems.

Additionally, the latest IoT technologies in conjunction with the most recent biodiversity algorithms and Artificial Intelligence models can be used to detect, classify, and eradicate specific weed species, as well as locate and identify fruits and vegetables, detect diseases, and boost ecosystem productivity without resorting to activities that harm the environment. Deep learning is already employed in several aspects of agriculture, but its application is still far from widespread. The most popular deep learning model in agriculture is the CNN. The adoption of novel techniques, such as attention mechanisms, new lightweight models, and single-stage detection models, can enhance the model's

performance. Performance metrics include accuracy, precision, recall, and F1-Score, and usually, precision, recall, and F1-Score are used together. The type of data that researchers employed the most is pre-existing datasets and from cameras.

In the future, crop management decision-support models for farmers may be created or enhanced to recommend the best course of action. Digital tools could be added that can instantly categorize weeds. The implementation of new sustainable practices backed by deep learning models and biodiversity monitoring will aid in managing the farm more efficiently and with less human labor.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tripathi, A.D.; Mishra, R.; Maurya, K.K.; Singh, R.B.; Wilson, D.W. Estimates for World Population and Global Food Availability for Global Health. In *The Role of Functional Food Security in Global Health*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 3–24. [CrossRef]
2. United Nations. Population. Available online: https://www.un.org/en/global-issues/population (accessed on 8 November 2022).
3. European Commission. *A Farm to Fork Strategy for a Fair, Healthy and Environmentally Friendly Food System. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM/2020/381 Final. Document 52020DC0381*; European Commission: Bruxels, Belgium, 2020.
4. Wang, A.; Zhang, W.; Wei, X. A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* **2019**, *158*, 226–240. [CrossRef]
5. United Nations. Water. Available online: https://www.un.org/en/global-issues/water (accessed on 8 November 2022).
6. Wato, M.A.T. The Agricultural Water Pollution and Its Minimization Strategies—A Review J. *Resour. Dev. Manag.* **2020**, *64*, 10–22. [CrossRef]
7. Hasan, A.S.M.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G.K. A survey of deep learning techniques for weed detection from images Comput. *Electron. Agric.* **2021**, *184*, 106067. [CrossRef]
8. 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. [CrossRef]
9. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sens.* **2020**, *12*, 3136. [CrossRef]
10. Littman, L.M.; Ajunwa, I.; Berger, G.; Boutilier, C.; Currie, M.; Doshi-Velez, F.; Hadfield, G.; Horowitz, M.C.; Isbell, C.; Kitano, H.; et al. *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report*; Stanford University: Stanford, CA, USA, 2021; Available online: http://ai100.stanford.edu/2021-report (accessed on 6 November 2022).
11. 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.
12. Espejo-Garcia, B.; Mylonas, N.; Athanasakos, L.; Fountas, S. Improving weeds identification with a repository of agricultural pre-trained deep neural networks. *Comput. Electron. Agric.* **2020**, *175*, 105593. [CrossRef]
13. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. *Information* **2019**, *10*, 349. [CrossRef]
14. 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. [CrossRef]

15. 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. [CrossRef]
16. 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. [CrossRef]
17. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
18. Sunasra, M. Performance Metrics for Classification Problems in Machine Learning. *Medium*. 11 November 2017. Available online: https://medium.com/@MohammedS/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b (accessed on 6 December 2022).
19. Javatpoint. Performance Metrics in Machine Learning. Available online: https://www.javatpoint.com/performance-metrics-in-machine-learning (accessed on 6 December 2022).
20. Swift, A.; Heale, R.; Twycross, A. What are sensitivity and specificity? *Evid. Based Nurs.* **2020**, *23*, 2–4. [CrossRef] [PubMed]
21. Rushikanjaria. Classification Model Performance Evaluation Using AUC-ROC and CAP Curves. *Geek Culture*. 5 July 2021. Available online: https://medium.com/geekculture/classification-model-performance-evaluation-using-auc-roc-and-cap-curves-66a1b3fc0480 (accessed on 7 December 2022).
22. 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. [CrossRef]
23. 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 (DASA), Sakheer, Bahrain, 8–9 November 2020; pp. 652–656. [CrossRef]
24. Azgomi, H.; Haredasht, F.R.; Motlagh, M.R.S. Diagnosis of some apple fruit diseases by using image processing and artificial neural network. *Food Control* **2023**, *145*, 109484. [CrossRef]
25. 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. [CrossRef]
26. 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. [CrossRef]
27. 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.
28. 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. [CrossRef]
29. 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. [CrossRef]
30. Kamath, R.; Balachandra, M.; Vardhan, A.; Maheshwari, U. Classification of paddy crop and weeds using semantic segmentation. *Cogent Eng.* **2022**, *9*, 2018791. [CrossRef]
31. 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. [CrossRef]
32. 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. [CrossRef]
33. 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. [CrossRef]
34. 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. [CrossRef]
35. 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. [CrossRef]
36. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo Algorithm Developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
37. 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)
38. Dyrmann, M.; Karstoft, H.; Midtiby, H.S. Plant species classification using deep convolutional neural network. *Biosyst. Eng.* **2016**, *151*, 72–80. [CrossRef]
39. 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. [CrossRef]
40. Gao, J.; Nuyttens, 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. [CrossRef]
41. 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. [CrossRef]
42. 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. [CrossRef]

43. 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. [CrossRef]

44. 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. [CrossRef]

45. Jin, X.; Che, J.; Chen, Y. Weed Identification Using Deep Learning and Image Processing in Vegetable Plantation. *IEEE Access* **2021**, *9*, 10940–10950. [CrossRef]

46. 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. [CrossRef]

47. 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. [CrossRef]

48. 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. [CrossRef]

49. 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. [CrossRef]

50. 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. [CrossRef] [PubMed]

51. 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. [CrossRef]

52. 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. [CrossRef]