

Review

Not peer-reviewed version

---

# Machine Learning for Precision Agriculture Using Imagery From Unmanned Aerial Vehicles (UAV): A Survey

---

[Imran Zualkernan](#)\*, [Diaa Addeen Abuhani](#), [Maya Hajj Hussain](#), Jowaria Khan, [Mohamed ElMohandes](#)

Posted Date: 3 May 2023

doi: 10.20944/preprints202305.0093.v1

Keywords: Precision Farming; UAVs; Agriculture; Machine Learning; Deep Learning; CNN; Transformers; GANs.



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

# Machine Learning for Precision Agriculture using Imagery from Unmanned Aerial Vehicles (UAV): A Survey

Imran Zualkernan <sup>1,\*</sup>, Diaan Addeen Abuhani <sup>1</sup>, Maya Haj Hussain <sup>1</sup>, Jowaria Khan <sup>1</sup> and Mohamed ElMohandes <sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering, American University of Sharjah, Sharjah, United Arab Emirates

\* Correspondence: izualkernan@aus.edu; +971-6-515-2953

**Abstract:** Unmanned Aerial Vehicles (UAV) are increasingly being used in a variety of domains and precision agriculture is no exception. Precision agriculture is the future of agriculture and will play a key role in long-term sustainability of agricultural practices. This paper presents a survey of how image data collected using UAVs has been used in conjunction with machine learning techniques to support precision agriculture. Numerous agricultural applications including classification of crop types and trees, crops detection, weed detection, cropland cover, and segmentation of farming fields are discussed. A variety of supervised, semi-supervised and unsupervised machine learning techniques for image-based precision agriculture are compared. The survey showed that for traditional machine learning approaches, Random Forests performed better than Support Vector Machines (SVM) and K-Nearest Neighbor Algorithm (KNN) for crop/weed classification. And, while Convolutional Neural Networks (CNN) have been used extensively, the U-Net-based models outperformed conventional CNN models for classification and segmentation tasks. Among the Single Stage Detectors (SSD), YOLO series performed relatively well. Two-Stage Detectors like R-CNN, FPN, and Mask R-CNN generally tended to outperform SSDs. Vision Transformers (ViT) showed promising results amongst transformer-based models which did not generally perform better than CNNs. Finally, Generative Adversarial Networks (GANs) have been used to address the problem of smaller datasets and unbalanced data.

**Keywords:** precision farming; UAVs; agriculture; machine learning; deep learning; CNN; transformers; GANs

## 1. Introduction

The rapid growth of the world's population is putting growing demands on food production. A large percentage of world's population is facing food insecurity today [1]. According to the Food and Agriculture Organization of United Nations (FAO), the demand for food production will continue to rise, reaching a staggering 70% increase by the year 2050 [2]. Such high demand cannot be sustained using conventional farming practices today. The problem is exacerbated by continuously diminishing natural resources used as inputs for farming. Use of traditional farming practices have also played a significant role in environmental degradation, including spread of water and atmospheric pollutants [3], degradation and erosion of soil [4], evolution of pesticide-resistant pests, and endangerment of human health due to excessive use of pesticides and agricultural chemicals [5]. Hence, new sustainable food production methods that maximize yield and minimize environmental impact must be developed. Using technology to support agricultural processes can potentially help reach this goal [6].

Precision agriculture, as defined by the International Society of Precision Agriculture (ISPA), is an agricultural management strategy that relies on the use of technology and agricultural data to improve the quality, sustainability, and yield of agricultural production [7]. Precision agriculture uses a wide array of sensors and monitoring devices to measure farming parameters like vegetation greenness, water content, nutrient status, and soil health [8]. These metrics help farmers make better

decisions on how to manage their fields, reduce resource waste, and increase yield and production [9]. The use of precision agriculture also conserves farmers' time and reduces manual labor. For example, precision agriculture can replace farmers manually surveying and assessing vegetation in their fields which is a tiresome, time-consuming, and an error-prone task [10].

The use of Unmanned Aerial Vehicles (UAVs) in precision agriculture has grown rapidly in the last few years. This growth is a result of a UAVs' ability to gather large amounts of information quickly, which can then be used to guide and enhance agronomic decision-making. In the early days of precision agriculture, most image data from fields was collected using ground cameras either mounted on unmanned ground vehicles (UGVs), or fixed next to vegetation patches. Later, satellites were used for image capturing. More recently, UAVs are being used in precision agriculture because of their ability to capture images at lower altitudes, at different angles, and at a higher speed than that of satellites [10]. Finally, recently, agricultural datasets to have become available and hence it is now possible to build a variety of machine learning models using these datasets [12].

This paper provides a survey of current research on applying machine learning to UAV image data for precision agriculture. While precision agriculture can use a variety of sensors [13], this paper limits itself to those studies that primarily used UAV image data. There are many previous survey papers related to this topic. For example, Kamalais et al. [14] covered various topics including disease detection, landcover classification, crop type classification, plant recognition, and fruit counting. The solutions discussed included various types of Convolutional Neural Networks (CNN) backbones like AlexNet and variations of VGG16. Since this paper appeared in 2018, many new deep learning models and architectures introduced after 2018 are not covered in this paper. Ren et al. [15] contained comparable material but also discussed newer models and backbones like YOLO detectors and fine-tuned AlexNet models. This paper classified models based on the backbone being used but did not cover a wider variety of problems requiring newer architectures. Meshram et al. [16] only focused on problems regarding disease and pest detection. Radoglou-Grammatikis et al. [17] provided a broad, non-technical overview of UAVs for precision agriculture with an emphasis on applications. Finally, Aslan et al. [18] is a recent survey paper that covered the use of UAV in both indoor and outdoor spaces but did so primarily in a non-technical manner.

The contributions of this paper are as follow:

- The paper discusses a wide variety of precision agriculture problems that can be addressed by using image data acquired from UAVs.
- The paper provides a technical discussion of the most recent papers using image data from UAVs to address agricultural problems.
- The paper evaluates the effectiveness of the various machine learning and deep learning techniques using image data to address agricultural problems using UAV data.
- The paper points out some fruitful future research directions based on the work done to date.

Rest of the paper is organized as follows. Key challenges with respect to precision agriculture are discussed first. Background on the types of images being collected using UAVs and evaluation techniques for models built using these techniques is then presented. Survey design is described next followed by a detailed presentation of the survey results. The paper ends with a discussion and a conclusion.

## 2. Challenges in Agriculture

Farmers face a variety of challenges many of which may be addressed using precision agriculture. This section briefly describes some agricultural challenges that can be addressed using precision agriculture utilizing image data from UAVs.

### 2.1. Plant Disease Detection and Diagnosis

The spread of plant diseases and the resulting loss in crop yields is a key problem for many farmers. According to the Food and Agriculture Organization of the United Nations (FAO), plant diseases have been the cause of around \$220 billion annual loss to the global economy [19]. To control the spread of diseases, and hence the yield loss, farmers must detect plant diseases at an early stage

and then take appropriate remedial measures. To detect diseases, farmers survey their fields to find obvious signs of infection. Once found, samples are taken from the infected plants to be observed under a microscope or similar instruments in the laboratory for a more reliable diagnosis [20]. Various laboratory assessments can also be performed for disease identification including Polymerase Chain Reaction (PCR) and Fluorescence in-situ Hybridization (FISH). Using such methods is laborious and time-consuming and requires access to experts in plant pathology. In addition, since these methods require plant sampling, they are damaging in nature [21].

## *2.2. Pest Detection and Control*

Pests like insects and weeds can result in significant crop yield loss. Insects cause yield loss by feeding on plants or by spreading plant diseases [22]. Weeds cause yield loss by consuming crop-growth resources including water and nutrients [23]. Detecting and curbing the spread of pests by applying pest control mechanisms is essential for ensuring a good yield. Like plant disease detection, traditional pest detection is a manual time-consuming and expert-reliant process. Once detected, a control mechanism must be applied to halt the spread of pests. One popular method of pest control is spraying a field with agrochemicals where herbicides are sprayed to eradicate weeds, and insecticides are sprayed to control insects. Traditionally, agrochemicals are not sprayed on the affected areas of the fields only. Rather, chemicals are applied across a field because detecting specific areas requiring agrochemicals is time-consuming [24]. As a result, current pest control methods are wasteful and unnecessarily expensive. Moreover, research has shown dangers of agrochemical use on the environment and human health. For example, pesticides have been found to pollute water and air and cause significant changes to soil ecosystems by causing harm to soil microorganisms. In addition, pesticide use has been found to cause adverse health effects, including weakening immunity and causing cancer [25]. Excessive use of pesticides also results in the cultivation of pesticide-resistant populations of pests. Currently, more than 500 species of insects are resistant to insecticides and around 270 species of weeds are resistant to herbicides [26]. Use of pesticides will become an ineffective pest control mechanism if the farmers continue to use traditional pesticide application methods. Excessive use, however, is not the only problem that makes traditional methods of pesticide use inefficient and ineffective. For pesticides to be most effective, correct identification of pests must be done to select the right type of pesticides. Subsequently, an accurate estimation of pests' vulnerability phase timings must be made to apply pesticides. Accuracy in time estimation is also required because early or late application of pesticides has little to no effect on the mitigation of pest spread. In order to choose the correct pesticides to use, and to determine the timing of application, an entomologist must survey the field, identify the pest types in the field, and predict the timing of their vulnerability phase [27]. Hence, for traditional pesticide-reliant pest control mechanisms to work, farmers also require expert help. Taking everything into account, traditional pest detection and control methods are not sustainable, especially with the growing demand for crop production.

## *2.3. Urban Vegetation Classification*

Urban vegetation plays a vital role in facing global climate change challenges. The dominance of a single type of tree results in rapid temporal changes in the ecosystem functions like carbon storage [28]. Machine learning can be used to classify different tree species within heterogeneous urban environments. UAV imagery with spectral information allows more accurate classification results that can be used later to create a better distribution of plants on the landscape. Further, improving the mapping capacities in spatial, spectral, and geometric domains in agriculture enables better analysis of urban landscapes and efficient resolutions to encounter the increasing thermal changes [29].

## *2.4. Crop Yield Estimation*

Accurate crop yield estimation helps create realistic plans for labor employment and agricultural produce storage [30]. Yield estimation is also important for making changes to crop management

practices to improve the final crop yield [31]. Traditionally, crop yield is estimated by finding the yield in a small sample area of a field and then generalizing the results to the entire field's area [32]. While seeming simple enough, in addition to being inaccurate, this method requires time-consuming manual work [33]. Manual crop counting is also an inefficient method of crop yield estimation with larger fields and more varied crop types. In addition, an obvious drawback of this method is that the estimation inaccuracies can lead to suboptimal plans for crop yield, labor, and storage.

### *2.5. Over and Under Irrigation*

Postharvest quality of crops depends on preharvest practices [34]. Appropriate irrigation is one of these preharvest practices that plays a crucial role in determining crop quality. Several crops are not drought resistant and, therefore, yields decrease considerably after short periods of water deficiency during production. For example, a study conducted by Mitchell et al. [35] found that deficit irrigation reduced fruit water accumulation and fresh fruit yield. In addition, Atay et al. [36] hypothesized that over-irrigation could have a negative impact on total yield and fruit quality. Finally, with water being a scarce resource in most production areas, efficient water management scheme that maintains crop yield but has a moderate and controlled level of moisture stress on their crops is required [29]. Multispectral images acquired from a UAV for water irrigation level recognition can potentially be used to help address over and under irrigation [37].

### *2.6. Seed Quality and Germination*

In crop planting, the germination rate and germination potential of seeds are important measures of their quality. Seed germination is the most critical stage of crop growth, and includes complex interactions such as water absorption, membrane reorganization, metabolic reorganization, and cell expansion. Selecting seeds with high germination rate can increase crop yields and help mitigate the risk of poor seed germination. However, during seedling stage, real-time monitoring, and detection of seed germination crucial for testing the quality of seeds, crop field management, and yield estimation. Owing to the low efficiency of traditional manual seedling rate monitoring, survey methods must gradually be replaced by UAVs and real-time video counting models.

### *2.7. Soil Quality and Composition*

Soil quality and composition are critical for maximizing the crops output and for increasing yield. The potential root zone in the soil should be well-tilled and fertilized with the needed minerals [34]. Balanced levels of nitrogen, water, and calcium improves the crops quality and reduce the post-harvest decay and vice versa. UAVs equipped with multispectral cameras may detect useful geospatial data such as water stress, nitrogen level, and other existing supplements [38]. Appropriate soil treatments can then be performed at the right time through foliar sprays.

### *2.8. Fertilizer Usage*

Use of fertilizers increases the yield of crops by providing plants the nutrients necessary accelerate growth. The type of fertilizer used depends on many factors including the crop type, the quality required, the purpose of use, and the diseases prevalent among the crop type. Under-use of fertilizers usually results in a reducing the quality of the crop because it may lower the sugar content and reduce the crops firmness. Meanwhile, over-fertilizing may result in multiple quality traits impairments such as total soluble solids, glucose, fructose, and pH issues [39]. Hence, a balanced fertilization level is necessary for crops. UAVs can be used to spray crops leaves or root soil with different combinations of nutrients needed in an effective and controlled manner to enhance the crops quality and its resistance to bacterial infections.

### *2.9. Quality of the Crop Output*

Farmers aspire to produce crops with the highest possible quality by designing quality-ensuring pre-harvesting, harvesting, and post-harvesting plans tailored to the crop. In the pre-harvest stage,

farmers are concerned with designing irrigation, fertilization, pesticide mitigation, and crop drainage plans that produce crops with required quality traits [40]. With these plans, crops are watered with a proper irrigation schedule, fertilized with the correct type of fertilizer and with the correct amount, and sprayed with the right kind and amount of pesticides [39], [40]. The harvesting stage also plays a role in the quality of the crop output. Farmers must harvest their crops during the correct time window to ensure that they adhere to their expected color, size, taste, and maturity characteristics. Traditionally, this is done using a variety of tests, including color, size, firmness, and acidity measurement. Lastly, in the post-harvest stage, farmers must design storage plans that ensure the quality of their harvest. Such plans must regard such important crop-specific factors as the harvested crops' storage time, and temperature and humidity requirements. Two examples of quality-ensuring storage plans include Dynamic Controlled Atmosphere (DCA) Storage and Heat Treatments [34]. Since crop pre-harvesting, harvesting, and post-harvesting plans require attention to crop types and careful assessment of their needs, farmers can leverage computer vision technology to perform these assessments and produce optimal plans, thereby reducing the need for manual labor and increasing the quality of the yield.

Table 1 below shows a summary of challenges in agriculture that potentially lend themselves to being addressed using image data from UAVs. As the table shows, UAVs can be potentially used to address all stages of the agriculture cycle.

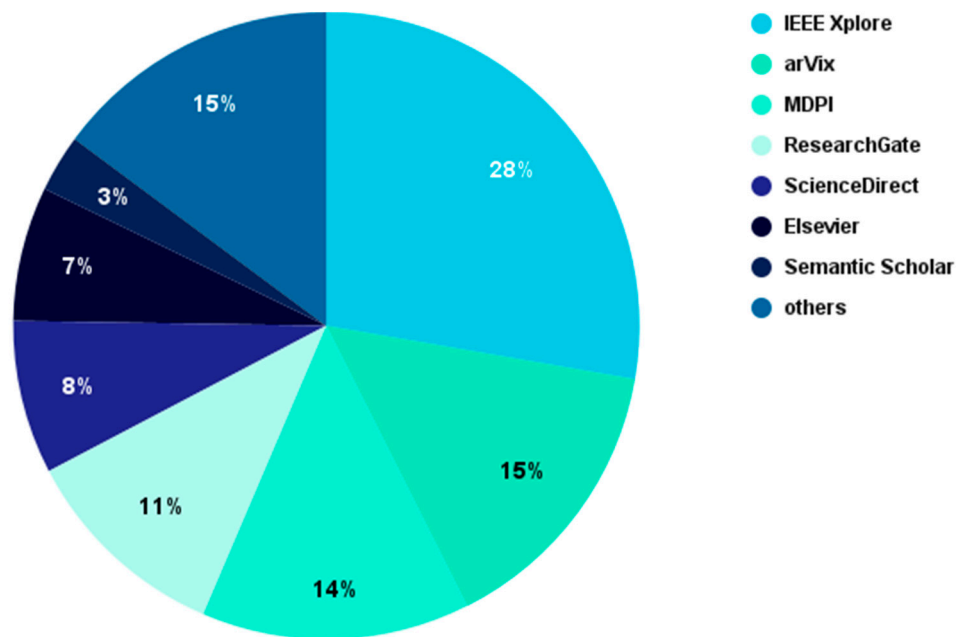
**Table 1.** Summary of Challenges in Agriculture.

Stage	Challenges
Pre-Harvesting	Disease Detection and Diagnosis, Seed Quality, Fertilisers Application, Field Segmentation, Urban Vegetation Classification
Harvesting	Crops Detection and Classification, Pest Detection and Control, Crop Yield Estimation, Tree counting, Maturity Level, Cropland Extent
Post-Harvesting	Fruit Grading, Quality retaining processes, Storage Environmental Conditions, Chemicals Usage detection

3. Survey Design

Journal articles and conference papers were collected using IEEE Xplore, arXiv, MDPI, ResearchGate, and ScienceDirect. "Deep learning," "Precision Farming," and "Agriculture" were the primary search terms utilized. The keywords "crops" and "segmentation" were added to other searches. From the search results, only articles published between 2017 and 2023 were included, and when applicable, the results were sorted by relevance and citation count. Figure 1 shows the distribution of resulting publications. The exclusion and inclusion of research articles was decided first by a preliminary abstract analysis, then by a full review of the article. The research papers which did not utilize UAV/aerial image datasets were excluded. In addition, the primary inclusion criteria for our research were as follows:

- The study must include a clear report on the performance of the models.
- The study must present an in-depth description of the model architecture.
- The study carries out detection/classification/segmentation tasks or a combination of these using UAV image datasets.



**Figure 1.** Sources Distribution of this Survey Paper.

The exclusion criteria of our research were as follows:

- The study is not indexed in a reputable database.
- The study does not propose any significant addition or change to previously existing deep learning or machine learning solutions in its domain.
- The study presents vague descriptions of the experimentation and classification results.
- The study proposes irrelevant or unsatisfactory results.

Based on these criteria, 70 papers were chosen for the survey. The papers were studied carefully to address the following major research questions:

- What data sources and image datasets were used in the paper?
- What type of pre-processing, data cleaning and augmentation methods were utilized?
- What type of machine learning or deep learning architectures were used?
- What overall performance was achieved, and which metrics were used to report the performance?
- Which architectures and techniques performed best for a class of agricultural problems?

## 4. Background

### 4.1. Image Data from UAVs

Raw images from a UAV are typically first corrected for displacements and distortions caused by terrain relief, camera tilt, etc. to create orthophoto images. The resulting images can be normal RGB images that define red, green, and blue color components for each individual pixel in an image. The RGB images are acquired using the standard visible light-sensitive cameras that usually only give surface-level information about the target data [41]. In addition to traditional imaging, UAVs for agricultural applications use multispectral images that capture different wavelength ranges across the electromagnetic spectrum. Multispectral data can be used to assess variations in plant/crop health that may be useful information for early treatment. Deep learning models using multispectral imaging have been developed [42]. The Near Infrared (NIR) spectral band images are acquired at 750-900 nm wavelength band and are primarily used for vegetation applications. NIR imaging provides additional beyond surface level information about the target data [43]. In such images, the red edge refers to the region in NIR range where rapid change in reflectance of vegetation is observed [44]. Similarly, Color-infrared (CIR) imagery also uses a portion from the NIR range. The invisible NIR light of CIR can be seen by the human eye by shifting it and the primary colors over. On CIR

imagery vegetation appears red, water generally appears black and urban structures like buildings and roads show as a light blue-green tint [45]. Finally, Normalized Difference Vegetation Index (NDVI) quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). The NDVI value ranges between -1 and 1 [46]. The NIR and Red channels are used for NDVI calculation as shown in Eq (1).

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

#### 4.2. Image Features Used in UAV Data

Many machine learning models used derived features from images acquired from a UAV. Examples of commonly used features include Hue-Saturation-Value (HSV) channels, Vegetation Indices (VI)s from RGBs such as Excess Green Index (ExG), excess green minus (ExGR), and the color index of vegetation extraction (CIVE). Other VIs are crop sensitive and can be derived from NIR spectrums such as NDVI, Ratio vegetation index (RVI), and Perpendicular vegetation index (PVI). Edge detectors could also be useful and are commonly used such as Gaussian, Laplacian, and Canny Filters [47], Gabor Filters, Gray Level Co-occurrence Matrix (GLCM) [48], geometric and statistical features [49] are among other useful features used in precision farming.

#### 4.3. Vision Tasks Using UAV Data

Precision agriculture applications using UAV image data are based on several computer vision tasks [50]. Image classification is the task of identifying which class the image of an object belongs to. Identification of weeds, for example, can be treated as an image classification task; given the image of a weed, identifying the image as a weed or non-weed. Another example is using image classification to identify different crops [51]. Image classification generally does not require isolating a particular object (e.g., a weed), but is based on observing general features in an image.

Object detection [52] is a related vision task that consists of identifying the location and labels of objects in an image. This task involves creating bounding boxes around objects and then labelling them. For example, for weed counting, one can detect all the weeds in an image and draw bounding boxes around them.

Another vision task is semantic segmentation that tries to identify objects that look similar or different from each other (e.g., weeds, ground, crop) at the pixel level [53]. For example, Zhang et al. [54] used segmentation to label pixels corresponding to purple rapeseeds to detect Nitrogen stress using UAV RGB data.

Finally, the instance segmentation task combines semantics segmentation and object detection to not only create a bounding box around an object, but to also then label each of the pixels of the object to belong to that specific instance. For example, in addition to identifying a weed, instance segmentation would also label each of the pixels of the weed and hence identifying shape of the weed as well.

#### 4.4. Evaluation Metrics

Several evaluation metrics have been used to assess and compare the machine learning methods used for the various vision tasks described earlier. This section provides a brief explanation of the most used metrics. Any additional metrics used by a paper's author/s are explained in the summary of the respective paper.

1. Accuracy as shown in Eq (2) is a measure of an algorithm's ability to make correct predictions. Accuracy is described as the ratio of the sum of True Positive ( $T_P$ ) and True Negative ( $T_N$ ) predictions to the algorithm's total number of predictions including false predictions ( $F_P + F_N$ ).

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (2)$$

- Precision as shown in Eq(3) is a measure of an algorithm's ability to make correct positive predictions. Precision is described as the ratio of True Positive ( $T_P$ ) predictions to the sum of True Positive ( $T_P$ ) and False Positive ( $F_P$ ) predictions.

$$Precision = \frac{T_P}{T_P + F_P} \quad (3)$$

- Recall as shown in Eq(4) measures an algorithm's ability to identify positive samples. Recall is the ratio of True Positive ( $T_P$ ) predictions made by the algorithm to the sum of its True Positive ( $T_P$ ) and False Negative ( $F_N$ ) predictions.

$$Recall = \frac{T_P}{T_P + F_N} \quad (4)$$

- F1-score as shown in Eq(5) is the harmonic mean of precision and recall. A high algorithm F1-score value indicates high accuracy. F1-score is calculated as follows:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

- Area Under the Curve (AUC) is the area under an ROC curve which is a plot of an algorithm's True Positive rate (TPR) Eq(6) Vs. its False Positive rate (FPR) Eq(7). An algorithm's True Positive Rate can be defined as the ratio of positive samples an algorithm correctly classifies to the total actual positive samples. The false positive rate, on the other hand, can be defined as the ratio of an algorithm's false positive sample classifications to the total actual negative samples.

$$TPR = \frac{T_P}{T_P + F_N} \quad (6)$$

$$FPR = \frac{F_P}{F_P + T_N} \quad (7)$$

- Intersection over Union (IoU) as shown in Eq(8) is an evaluation metric used to assess how accurate a detection algorithm's output bounding boxes around an object of interest in an image (e.g., a weed) are compared to the ground truth boxes. IoU is the ratio of the intersection area between a bounding box and its associated ground truth box to their area of union.

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ intersection} \quad (8)$$

- Mean Average Precision (mAP) as shown in Eq (9) is used to assess the quality of object detection models. This metric requires finding a model's average AP across its classes. The calculation of AP requires calculating a model's Precision and Recall, followed by drawing its precision-recall curve, and finally, finding the area under the curve.

$$AP = \int_0^1 P(R) dR \quad mAP = \frac{1}{n} \sum_{k=1}^n AP_k \quad (9)$$

- Average residual as shown in Eq(10) is used to assess how erroneous a model is. Average residual displays the average difference between a model's predictions and ground truth values.

$$Average\ Residual = \frac{1}{n} \sum_1^n prediction - real\_value \quad (10)$$

- Root Mean Square Error (RMSE) as shown in Eq (11) is used to assess an algorithm's ability to produce numeric predictions that are close to ground truth values. RMSE is calculated by finding the square root of the average distance between an algorithm's predictions and their associated truth values.

$$RMSE = \sqrt{\frac{\sum_1^n (prediction - real\_value)^2}{n}} \quad (11)$$

- Mean Absolute Error (MAE) as shown in Eq (12) is an error metric used to assess how far off an algorithm's numeric predictions are from truth values. MAE is calculated by finding the average value of the absolute difference between predictions and truth values.

$$MAE = \frac{\sum_1^n |prediction - real\_value|}{n}$$

(12)

11. Frames Per Second (FPS) is a measure used to assess how fast a machine learning model is at analyzing and processing images.

5. Survey Results

70 papers were shortlisted based on the filtering methodology described earlier. These papers used a variety of cameras to capture images from UAVs. The cameras used ranged from low-end cameras like the Raspberry Pi NoIR Camera Module to commercial UAV cameras like the DJI FC6310 with high resolution imaging capabilities. In addition, consumer cameras like the 36.4-megapixel Sony A7R (RGB) camera were also deployed on larger UAVs. As Figure 2 shows, a wide image resolution was used in the papers reviewed here. The size of images ranged from 64x64 pixels to 7000x5000 pixels.

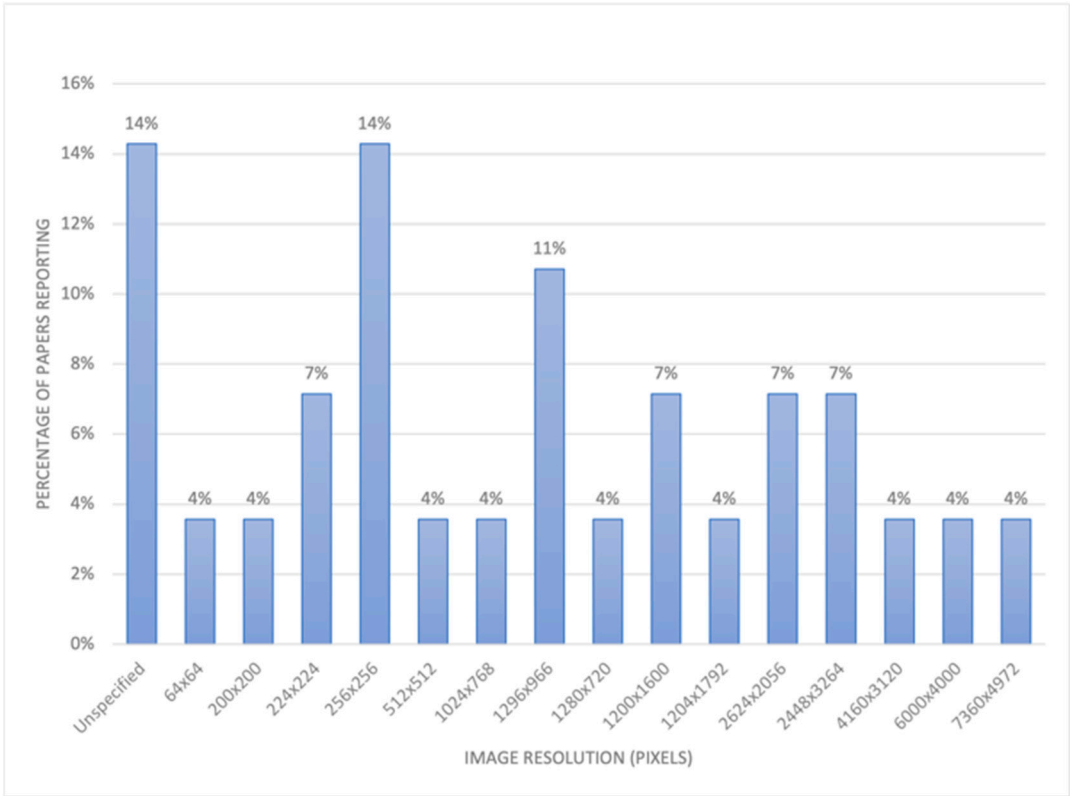


Figure 2. Image size (pixels) Used for Vision Tasks using UAV Data.

Figure 3 shows the percentage of papers addressing the various precision agricultural issues using the taxonomy of Liliane and Charles [55]. As Figure 3 shows, spatial segmentation and pesticide and disease treatment were the primary areas of interest in the reviewed papers.

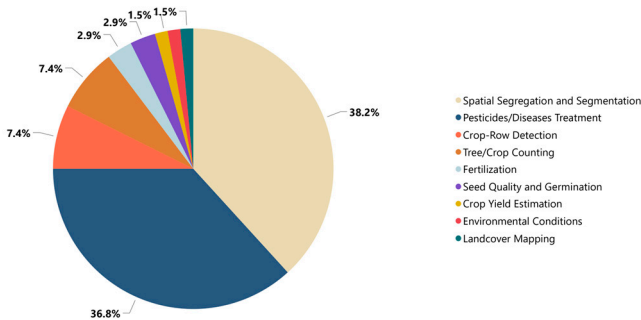


Figure 3. Various Precision Agricultural Issues Addressed in the Papers.

Results of the filtered papers are organized according to the various machine learning techniques used. Papers using traditional machine learning techniques are discussed first followed by those utilizing neural networks and deep learning methods.

### 5.1. Traditional Machine Learning

#### 5.1.1. Support Vector Machines (SVM)

Support Vector Machines (SVM) were used for classifying vegetation by health status [47], classifying trees by type [56], identifying and classifying weeds to generate weed maps [57], and lastly, to segment crop rows [48].

Tendolkar et al. [47] proposed the use of Agrocopter, a multipurpose farming drone to assess and evaluate plant health status, and for taking corrective actions. The system assessed plant health based on the NVDI index, texture, and color features of the individual pixels. These features were extracted utilizing a filter bank of 17 Gaussian and Laplacian filters. SVM was then used to perform semantic segmentation on the image pixels, and to classify the pixels as healthy or unhealthy. Finally, a segmented mask was generated and used to find the health ratio of the images based on the ratio of the area of healthy pixels to the total area of the image. The health ratio was then used to classify images into healthy, moderately healthy, and unhealthy. The trained model had 85% precision, 81% recall, and an F1-score of 79%.

Natividade et al. [56] proposed a pattern recognition system (PRS) to identify and classify vegetation using NDVI scale as a segmentation threshold. An SVM was trained on two data sets: a tree dataset with five classes and a vineyard dataset with three classes. The best models achieved an accuracy of around 72% on the two data sets.

Pérez-Ortiz et al. [57] introduced a UAV-based weed mapping system for the early detection of weeds in crop fields. They used a semi-supervised SVM (SSVM). The system used crop-row detection, vegetation indices, and spectral features to classify pixels in field images as belonging to one of three classes of crop, weed, or soil. Crop row detection was introduced to improve classifier performance in differentiating crops and weeds because their spectral features were similar. The proposed system took UAV-captured images, partitioned them into 1000x1000 pixel images, and then calculated the vegetation index of all image pixels. NDVI was used for multispectral images and the Excess Green index (ExG) was employed for visible images. The Otsu thresholding procedure was then applied to the vegetation indices to create thresholds that divided the indices into three classes where the highest Vegetation Index (VI) pertaining to crops, lower values to weeds, and the lowest values to soil. The image was then binarized by taking crop pixels as 1's and weed and soil pixels as 0's. The binarized image was then fed into the Hough Transform (HT) method to detect crop rows in the images. Finally, a crop row data feature, along with VI and spectral features, was used to train different machine learning models to classify pixels as soil, crop, or weed. The SSVM returned a MAE of 12.68%.

César Pereira et al. [48] compared the performance of multiple machine learning algorithms for the problem of crop row segmentation. Their study used a single image of a sugar cane field as its dataset and compared the segmentation results of running this image through different classifiers to a manually labelled image. The manually segmented image's pixels were classified into the two classes of crop row and the background. Spectral features were extracted using ExG and VI, and textural features were extracted through a four-filter Gabor filter bank and a Gray Level Co-Occurrence Matrix (GLCM). The feature vectors and color features (RGB) were used to train SVM models. For the Linear SVM model, the best combination of features was RGB, EXG, and Gabor filters. This combination yielded an F1-Score of 88.01% and an IoU percentage of 78.86%. The worst feature combination was RGB and GLCM. This combination yielded an F1-Score of 62.48% and an IoU percentage of 46.08%.

Table 2 shows a summary of the use of SVM with drone image data. Some similarities can be observed. Many (e.g., [47],[56],[57]) used the NDVI scale to perform pixel-wise classification. While

others (e.g., [57] and [48]) used the ExG index for classification. Both [56] and [48] used a Radial Basis Function kernel to find the optimal hyperplane for separating the dataset classes.

**Table 2.** Support Vector Machine Summary.

Paper	SVM			
	Model / Architecture	Approach	Comments	Best Results
Tendolkar et al. [47]	SVM	Dual-step based approach of pixel-wise NDVI calculation and semantic segmentation helps in overcoming NDVI issues.	Model was not compared to any other.	Precision: 85% Recall: 81% F1-score: 79%
Natividade et al. [56]	SVM	Pattern recognition system allows for classification of images taken by low-cost cameras.	Accuracy, precision, and recall values of model vary highly between datasets	Dataset #1 (1st configuration): [Accuracy: 78%, Precision: 93%, Recall: 86%, Accuracy: 72%] Dataset #2: [Accuracy: 83%, Precision: 97%, Recall: 94%, Accuracy: 73%]
Pérez-Ortiz et al. [57]	SVM	Able to detect weeds outside and within crop rows. Does not require a big training dataset	Segmentation process produces salt and pepper noise effect on images. Training images are manually selected. Model inference is influenced by training image selection.	Mean Average Error (MAE): 12.68%
César Pereira et al. [48]	LSVM	Model can be trained fast with a small training set	Image dataset is small and simple, containing images of sugarcane cultures only.	Using RGB + EXG + GABOR filters: IOU: 0.788654 F1: 0.880129

5.1.2. K-Nearest Neighbors (KNN)

The K-nearest neighbor algorithm (KNN) has been used extensively in precision agriculture in land cover classification [58], sugarcane planting lines detection/fault study [59], and crop row segmentation [48].

Rodríguez-Garlito and Paz-Gallardo [58] proposed a KNN based land cover classification system. This system classified land cover into olive trees, soil, weeds, and shadow. In this system, high-resolution, multi-spectral images of the studied field were first captured using a UAV. These images went through spatial partitioning to reduce memory costs of the machine learning algorithm. As a result, processing windows were formed, with each window holding the spectral information of a row of image pixels. The KNN algorithm was then applied to one processing window at a time to perform land cover classification, and to classify individual pixels into the classes to which they belonged. The trained KNN model had a precision of 95.5%, an accuracy of 91.8%, and a balanced accuracy score of 90.9%. Similarly, Rocha et al. [59] used KNN to detect gaps in curved sugarcane planting lines from aerial images. The training and test sets were created using RGB images and classified based on Decision Tree, Linear Discriminant Analysis, and KNN. KNN had the best results with a relative error of 1.65% and effectively evaluated the planting conditions.

Pereira Júnior et al. [48] studied the use of the KNN algorithm in crop row segmentation. Two KNN models with two different K values of 3 and 11 were used. Constructing a KNN model with a K value of either 3 or 11 yielded similar results. The models used the Euclidean distance and using RGB, ExG, and Gabor filters as features, both models achieved an IoU score of about 76% and an F1-score of about 86%. Results for applying K-NN are summarized in Table 3.

**Table 3.** K-Nearest Neighbors Summary.

Paper	KNN			
	Model/ Architecture	Strengths	Comments	Best Results
César Pereira et al. [48]	KNN3, KNN11	Simplest algorithms to implement amongst implemented algorithms in the paper	Models did not achieve the best results in the paper	KNN3 and 11 with RGB+ EXG+GABOR filters: [IOU: 0.76, F1: 0.86]
Rodríguez-Garlito and Paz-Gallardo [58]	KNN	Uses an automatic window processing method that allows for the use of ML algorithms on large multispectral images.	Model did not achieve the best results in the paper	Approximate values: AP: 0.955 Accuracy score: 0.918
Rocha et al. [59]	KNN	Best performing classifier	Model cannot perform sugarcane line detection and fault measurement on sugarcane fields of all growth stages	Relative Error: 1.65%

5.1.3. Decision Trees (DT) and Random Forests (RF)

Decision tree classifiers were used in precision agriculture to classify vegetation like trees and vineyards, [56]. Similarly, the random forest algorithm was used to classify sugar beet crops and weeds [49].

Natividade et al. [56] used decision trees to detect and classify trees and vineyards in a field, where trees were classified into five distinct types, and vineyards into three types. On the tree data set, the best model resulted in 87% precision, 88% recall, and 74% accuracy. On the vineyard data, 87% precision, 90% recall and 79% accuracy were achieved.

Lottes et al. [49] proposed a crop and weed detection, feature extraction, and classification system that could identify and classify sugar beets and several types of weeds. NDVI, and ExG were used as features. A segmented mask based on the VI threshold was then used to extract a spectral feature vector per segmented object in the image and a feature vector per key point in the image. These feature vectors, along with geometric and statistical features, were used to train a Random Forest model. PHANTOM and MATRICE-graining data sets contained UAV-captured images of crops and weeds while JAI training dataset contained ground-captured images. The PHANTOM dataset was used to test how well the model could classify vegetation into sugar beet crops, saltbush weeds, chamomile weeds, and other weeds. The model yielded a precision of 85% for both saltbush and chamomile weeds. The recall values were 95% and 87% for saltbush weeds and chamomile weeds, respectively. Lastly, a recall of only 45% was attained for other weeds. The overall accuracy of the model was 86%. When weed-type classification was ignored, and vegetation was classified into two classes, 99% recall and 97% precision was achieved. Table 4 summarized the two studies using decision trees.

**Table 4.** Decision Trees and Random Forests Summary.

Paper	DT & RF			
	Model/ Architecture	Strengths	Comments	Best Results
Natividade et al. [56]	DT	Pattern recognition system allows for classification of images taken by low-cost cameras.	Model did not outperform SVM on all chosen metrics	Dataset #2: [Accuracy: 77%, Precision: 87%, Recall: 90%, Accuracy: 79%]
Lottes et al. [49]	RF	Model can detect plants relying only on its shape.	Low precision and recall for detecting weeds under the “other weeds” class	Saltbush class recall: 95%, Chamomile class recall: 87%, sugar beet class recall: 78%, recall of other weeds class: 45% Overall model accuracy for predicted objects: 86%

5.2. Neural Networks and Deep Learning

5.2.1. Convolutional Neural Networks (CNN)

Convolutional neural networks (CNN) have been used extensively in analyzing images for precision agriculture. In specific, transfer learning has been often used successfully using a variety of pre-trained models including Inception V3, VGG, etc. For example, Crimaldi et al. [60] used Inception V3 model and achieved 78.1% accuracy for classifying a crop into one of 14 crop types using data consisting of 54,309 images. Milioto et al. [61] built a CNN model using RGB and NIR camera images. The model had 97.3% accuracy for images of early crop growth and 89.2% accuracy for images of crops in later stages. However, both models had the same recall percentage, with the early stage scoring 98% and the later stage scoring 99%. Similarly, Bah et al. [62] used the AlexNet model on spinach, beet, and beans datasets, and got a precision of 93%, 81%, and 69%, respectively. The authors claimed that the bad results were primarily due to leaves overlapping between crops and weeds. Reddy et al. [63] used a customized CNN model for their work on plant species identification and achieved 99.5% precision for Flavia, Swedish leaf, and UCI leaf datasets. Sembiring et al. [64] focused on tomato plant disease detection. Their proposed model achieved 97.15% validation accuracy using the tomato leaf dataset from Plant village. However, their model did not achieve the highest validation accuracy among all four trained models. The highest accuracy score of 98.28% was achieved by the VGG16 model. Geetharamani et al. [65] achieved a classification accuracy of 96.46% using a customized nine-layer CNN model. R. et al., [66] used a residual learning CNN with an attention mechanism. The goal was to perform real-time corn leaf disease recognition. They also used the Plant Village Disease Classification challenge dataset. An overall accuracy of 98% was achieved. Nanni et al. [67] used different combinations of CNNs, including ResNet50, GoogleNet, ShuffleNet, MobileNetv2, and DenseNet201, with different Adam optimization methods. These CNN models were trained on three datasets of insect images: The Deng dataset, the IP102 dataset, and the Xie2 dataset. The best performing CNN achieved state-of-the-art accuracy on both insect datasets: 95.52% on Deng, a score that competed with human expert classifications, and 73.46% on IP102.

Atila et al. [68] proposed using the EfficientNet architecture for plant disease classification on the plant village dataset and achieved 99.91% and 99.97% accuracy on original and augmented datasets, respectively. Prasad et al. [69] proposed a two-step machine learning approach that analyzed low-fidelity and high-fidelity images from drones in sequence, preserving efficiency as well as the accuracy of plant diagnosis. The pathology 2020 dataset and a set of synthetically generated images were used. A semi-supervised model derived from EfficientNet called EfficientDet was used. The end goal was to perform segmentation and classification. The model scored 75.5% for the average accuracy of the identifier model. Albattah et al. [70] proposed a customized model of using EfficientNet called EfficientNetV2-B4 backbones to address plant disease classification. The Plant

Village dataset and additional UAV images were used to train the model. The results were 99.63%, 99.93%, 99.99%, 99.78% for precision, recall, accuracy, and F1-score, respectively.

Mishra et al. [71] developed a standard CNN model to detect corn plant diseases in real-time. The model was deployed on an Intel Movidius NCS and a Raspberry Pi 3b+ module. The authors used the Plant-Village Disease Classification Challenge dataset and divided the images into three classes: Rust, Northern leaf blight, and Healthy. The system achieved an accuracy of 98.40% using a GPU and 88.56% on the NCS chip. Bah et al. [72] used unsupervised data labeling for weed detection from UAV images. The dataset consisted of two fields: Beans and Spinach. Each dataset was divided into the two classes or crop and weed. Two-thirds of the data was labelled in a supervised manner, while the third was labelled using unsupervised methods. The ResNet-18 model was used to perform the classification. The ResNet18 significantly outperformed SVM and RF methods in the bean field as it achieved an average AUC of 91.7% on both supervised and unsupervised labelled data in comparison to 52.68% using SVM and 66.7% using RF. On the other hand, RF resulted in a slightly better average AUC% in the Spinach field to that achieved using ResNet-18.

Zheng et al. [73] proposed multiple CNN models to estimate percent canopy cover as well as vineyard leaf area index in each field. The authors compared the estimation performance of five different models, including a CNN-ConvLSTM model, a Vision Transformer model, a Joint Model, an Xception model, and a ResNet-50 model. The five models were trained on a dataset containing approximately 840 images extracted from UAV videos taken of vineyard fields at Alcorn State University. The five models were evaluated using the RMSE of both leaf area index (LAI) and Percent Canopy Cover. For the prediction of leaf area index, Xception, CNN-ConvLSTM, Vision Transformer, ResNet-50, and Joint model had RMSE of 0.28, 0.32, 0.34, 0.41 and 0.43 respectively. For predicting percent canopy cover, Xception, CNN-ConvLSTM, Vision Transformer, ResNet-50, and Joint model had RMSE of the 4.01, 4.50, 4.56, 5.98, and 6.08 respectively. Clearly, Xception performed best for both LAI estimation and percent canopy cover estimation.

Yang et al. [74] proposed a method of multi-source data fusion for disease and pest detection of grape foliage using the ShuffleNet V2 model. The dataset consisted of 834 groups of grape foliage images. Each group contained three types of images of grape foliage: RGB Image (RGBI) (2592 × 1944, 3 channels), Multispectral Image (MSI) (409 × 216, 25 channels), and Thermal Infrared Image (TIRI) (640 × 512, 3 channels). The accuracy of MSI was 82.4%, RGB was 93.41%, and the TIRI was 68.26%.

Briechele et al. [75] used multispectral images to classify tree species and standing dead trees. They used the PointNet++ model. The data used was UAV-based Light Detection and Ranging including Laser Echo Pulse Width (LIDAR) data and five-channel MS imagery. They also applied segmentation to the images during the preprocessing of the data. Their model achieved an accuracy of 90.2%.

Aiger et al. [76] proposed a method of image classification based on multi-view image projections. Their method's used projections of multiple images at multiple depth planes near the reconstructed surface. This enabled the classification of categories whose most noticeable aspect was appearance change under different viewpoints, such as water, trees, and other materials with complex reflection/light response properties. They obtained the best accuracy of 96.3% on their proposed 3D CNN.

**Table 5.** Convolutional Neural Networks Summary.

Paper	CNN			
	Model/ Architecture	Strengths	Comments	Best Results
Crimaldi et al. [60]	Inception V3	The identification time is 200ms which is good for real-time applications	Low accuracy	Accuracy of 78.1%
Milioto et al. [61]	CNN model fed with RGB+NIR camera images	High accuracy for early growth stage	Low accuracy for the later growth stage	Early growth stage Accuracy: 97.3% Recall: 98%

Paper	CNN			
	<i>Model/ Architecture</i>	<i>Strengths</i>	<i>Comments</i>	<i>Best Results</i>
				Later growth stage Accuracy: 89.2% Recall: 99%
Bah et al. [62]	AlexNet	Less images with high resolution from a drone	Overlapping of the leaves between crops and weeds	Best precision was for the Spinach dataset with 93%
Reddy et al. [63]	Customized CNN	The results had a high precision and recall	Large dataset	Precision of 99.5% for the leaf snap dataset. The flavia, Swedish leaf, UCI leaf datasets had a recall of 98%.
Sembiring et al. [64]	Customized CNN	Low training time compared to other models compared in the paper	Not the highest performing model compared in the paper	Accuracy of 97.15%
Geetharamani et al. [65]	Deep CNN	Can classify 38 distinct classes of healthy and diseased plants	Large dataset	Classification accuracy of 96.46%
R. et al. [66]	Residual learning CNN with attention mechanism	Prominent level of accuracy and only 600k parameters which is lower than the other papers compared in this paper	Large dataset	Overall Accuracy of 98%
Nanni et al. [67]	ensembles of CNNs based on different topologies (ResNet50, GoogleNet, ShuffleNet, MobileNetv2, and DenseNet201)	Using Adam helps in decreasing the learning rate of parameters whose gradient changes more frequently	IP102 is a large dataset	95.52% on Deng and 73.46% on IP102 datasets
Bah et al. [77]	CrowNet	Able to detect rows in images of several types of crops	Not a single CNN model	Accuracy: 93.58% IoU: 70%
Atila et al. [68]	EfficientNet	Reduces the calculations by the square of the kernel size	Did not have the lowest training time compared to the other models in the paper	Plant village dataset Accuracy: 99.91% Precision: 98.42% Original and augmented datasets Accuracy: 99.97% Precision: 99.39%
Prasad et al. [69]	EfficientDet	Scaling ability and FLOP reduction	Performs well for limited labelled datasets however, the accuracy is still low	Identifier model average accuracy: 75.5%
Albattah et al. [70]	EfficientNetV2-B4	Really reliable results and has low time complexity	Large dataset	Precision: 99.63% Recall: 99.93% Accuracy: 99.99% F1: 99.78%
Mishra et al. [71]	Standard CNN	Can run on devices like raspberry-pi or smartphones and drones. Works in real-time with no internet.	NCS recognition accuracy is not good and can be improved according to the authors	Accuracy GPU: 98.40% NCS chip: 88.56%

Paper	CNN			
	Model/ Architecture	Strengths	Comments	Best Results
Bah et al. [72]	ResNet18	Outperformed SVM and RF methods and uses unsupervised training dataset	Results of the ResNet18 are lower than SVM and RF in the spinach field	AUC: 91.7% on both supervised and unsupervised labelled data
Zheng et al. [73]	Multiple CNN models including: CNN- Joint Model, Xception model, and ResNet-50 model.	Compares multiple models	Joint Model had trouble with LAI estimation, and the vision transformer had trouble with percent canopy cover estimation.	Xception model: 0.28 CNN-ConvLSTM: 0.32 ResNet-50: 0.41
Yang et al. [74]	ShuffleNet V2	The total params were 3.785 M which makes it portable and easy to apply	Not the least number of params when compared to the models in the paper	Accuracy MSI: 82.4% RGB: 93.41% TIRI: 68.26%
Briechelet et al. [75]	PointNet++	Good score compared to the models mentioned in the paper	Not yet tested for practical use	Accuracy: 90.2%
Aiger et al. [76]	CNN	Large-Scale, Robust, and high accuracy	Low accuracy for 2D CNN.	96.3% Accuracy

### 5.2.2. U-Net Architecture

The U-Net architecture was originally introduced in the medical domain by Ronneberger et al. [78] and is commonly used for image segmentation. U-Net follows an encoder-decoder architecture. Many factors like density of the crops, flight height of the drone, and the growth stage have an impact on how well U-Net will perform. According to Kitano et al. [79] U-Net did not perform well when the plants were remarkably close together. However, some techniques could be used to solve this problem, such as using the opening morphological operator.

Lin et al. [80] used U-Net to achieve an accuracy of 95.5% and a RMSE of 2.5% with 1000 manually labelled training images. Arun et al. [24] achieved an accuracy of 95.34% and an RMSE of 7.45 using reduced U-Net by designing an efficient pixel-wise for weeds and crops in agricultural field images. Hoummaidi et al. [81] used the U-Net model to perform vegetation extraction and achieved an overall accuracy of 89.7%. However, Palm trees and Ghaf trees had higher detection rates of 96.03% and 94.54%, respectively. The authors justified their results with the fact that trees were obstructed by other trees. Palm trees also caused some errors due to their physical characteristics and the small crown sizes of some trees. The authors suggested that including young Palms in the training data could improve the crown size error rate. Doha et al. [82] used the U-Net architecture to detect crop rows by performing semantic segmentation on vertical aerial images. Zhang et al. [83] used the Dual flow U-Net (DF-U-Net) to detect yellow rust severity in farmlands. The dataset was from the Yangling experiment field, which used a RedEdge camera on board a DJI M100 UAV with a sensor size of  $1336 \times 2991$ . The F1-score, accuracy, and precision scores were 94.13%, 96.93%, and 94.02%, respectively. Sparse Channel Attention (SCA) was designed to increase the receptive field of the network and improve the ability to distinguish each category. Using U-Net, Lin et al. [80] achieved high accuracy with a small dataset. Similarly, with only 48 images, Tsuichihara et al. [84] achieved an accuracy of about 80% for detecting broad-leaved weeds. Table 6 provides a summary of studies using the U-Net architecture.

Table 6. U-Net Summary.

Paper	U-Net			
	Model/ Architecture	Strengths	Comments	Best Results
Lin et al. [80]	U-NET	Can detect overlapping sorghum panicles	The performance decreases with lower number of training images (<500)	Accuracy: 95.5% RMSE: 2.5%
Arun et al. [24]	Reduced U-NET	Reduces the total number of parameters and results in a lower error rate	The comparison was done with models that were used to problems not related to agriculture which is what this model was used for	Accuracy: 95.34% Error rate: 7.45%
Hoummaidi et al. [81]	U-NET	Real-time and the use of multispectral images	Trees obstruction and physical characteristics causes it to have errors, however it can be improved using a better dataset	Accuracy: 89.7% Palm trees Detection rate: 96.03% Ghaf trees Detection rate: 94.54%
Doha et al. [82]	U-NET	The method they used can refine the results of the U-Net to reduce errors as well as do frame interpolation of the input video stream	Not enough results were given	Variance: 0.0083
Zhang et al. [83]	DF-U-Net	Reduces the computation load by more than half and had the highest accuracy among other models compared	Early-stage rust disease is difficult to recognize	F1: 94.13% OA: 96.93% Precision: 94.02%
Tsuichihara et al. [84]	U-Net	Accuracy is 80% for only 48 images trained	Low accuracy and that is due to the small number of images which is due to manually painting 6 colors on each image	Accuracy: ~80%

5.2.3. Other Segmentation Models

Efficient Dense modules of Asymmetric Convolution (EDANet) is another model that works well for real-time semantic segmentation. Therefore, EDANet can be useful for real-time applications like UAVs. Yang et al. [85] proposed an EDANet that does semantic segmentation for detecting rice lodging. Lodging occurs when the stem weakens, and the plant falls over. EDANet outperformed many systems because of its efficiency, low computational cost, and model size. The model identified normal rice at 95.28% and lodging at 86.17% accuracy. The model accuracy was improved to 99.25% when less than 2.5% of rice lodging was neglected.

Weyler et al. [86] proposed ERFNet-based instance segmentation model that segments individual crop leaves in plant imagery to extract relevant phenotyping information and then groups the instances that belong to one crop together. This model made use of two decoders, one of which was used to predict the offset of image pixels from leaf regions, while the other was used to predict the offset of image pixels from plant regions. The two decoder outputs were then used to generate one image with leaf clusters and another with plant clusters. The model was trained on a dataset of 1,316 RGB images of sugar beet fields captured by a camera onboard a UAV. The model was evaluated on its ability to perform crop leaf segmentation as well as full crop segmentation. In crop leaf segmentation, the model was able to achieve an average precision of 48.7% and an average recall of 57.3%. The model achieved an average precision of 60.4% and an average recall of 68% for crop segmentation.

Guo et al. [87] developed a three stages model to perform plant disease identification for smart farming. The model located the diseased leaves using a Region Proposal Network (RPN) algorithm trained on a leaf dataset in complex environments, after which regression and classification neural networks were used to locate and retrieve the diseased leaves. Later, the Chan-Vese algorithm was used to perform segmentation based on the set zero level set and minimum energy function. Finally, the diseases were identified using a pre-trained transfer learning model. The proposed model outperformed the traditional ResNet-101 model significantly with an accuracy of 83.75% in comparison to 42.5% by the latter.

Sanchez et al. [88] used a multilayer perception (MLP) neural network for the early detection of broad-leaved weeds and grass weeds in wide-row crops from UAV imagery. The data was manually collected using a UAV quadcopter equipped with a low-cost RGB camera. Image segmentation was done using the multiresolution segmentation algorithm (MRSA). The model achieved an average overall accuracy of 80.9% on two classes of crops.

Zhang et al. [89] proposed a unified CNN called UniStemNet for joint crop recognition and stem detection in real-time. The architecture of UniStemNet is like Mask R-CNN. The architecture consists of a backbone and two subnets in which the first performs crop recognition, and the other performs stem detection simultaneously. The backbone consists of five convolutional stages, where the first is a standard CNN with batch normalization while the other four contain two MobileNet2 inverted residual modules (IRMs). The subnets follow a varied-span feature fusion structure as each has different detection targets. The evaluation was performed on the open-source CWF-788 dataset, and labels were manually annotated. The model obtained an F1-score of 97.4% and an IoU score of 94.5 in segmentation which was slightly lower than that achieved by CR-DSS [90]. Nonetheless, the model achieved the best-known results in stem detection with an SDR of 97.8%. Summary of other segmentation models is presented in Table 7 below.

**Table 7.** Other Segmentation Models Summary.

Paper	Other Segmentation Models			
	Model/ Architecture	Strengths	Comments	Best Results
Yang et al. [85]	EDANet	Improved prior work on identifying and lodging by 2.51% and 8.26% respectively	Drone images taken from a greater height do not perform well, however, with the method they proposed it can have reliable results	Identify rice Accuracy: 95.28% Lodging Accuracy: 86.17% If less than 2.5% lodging is neglected, then the accuracy increases to 99.25%.
Weyler et al. [86]	ERFNet-based instance segmentation	Data was gathered from real agricultural fields	Comparison was made on different datasets	Crop leaf segmentation Average precision: 48.7% Average recall: 57.3% Crop segmentation Average precision: 60.4% Average recall: 68%
Guo et al. [87]	Three stage model with RPN, Chan-Vese algorithms and a transfer learning model	Outperformed the traditional ResNet-101 which had an accuracy of 42.5% and is unsupervised	The Chan-Vese algorithm runs for a long time	Accuracy 83.75%
Sanchez et al. [88]	MLP	Carried in commercial fields and not under controlled conditions	The dataset was captured noon to avoid shadow	Overall accuracy on two classes of crops: 80.09%
Zhang et al. [89]	UniStemNet	Joint crop recognition and stem detection in real-time. Fast and can finish processing each image within 6ms.	The scores of this model were not always the best and the differences were small	Segmentation F1: 97.4% IoU: 94.5 Stem detection SDR: 97.8%

#### 5.2.4. YOLO ONLY LOOK ONCE (YOLO)

You Only Look Once (YOLO) is a real-time object detection neural network model where a single stage neural network is applied to the full image. The network divides the image into regions and predicts bounding boxes along with probabilities for each region. YOLO has been gaining popularity lately in for agricultural disease and crop detection. For example, Chen et al. [91] proposed a UAV to photograph and detect pests and employed a Tiny-YOLOv3 model built on NVIDIA Jetson TX2 to recognize their position in real-time. The detected pest positions could later be used to plan optimal pesticide spraying routes, which agricultural UAVs would later follow. The model attained the best mAP score of 95.33% and 89.72% on 640\*640 pixels test images.

Similarly, Qin et al. [92] proposed a solution for precision crop protection based on a light deep neural network (DNN) called Ag-YOLO consisting of a modified version of ShuffleNet-v2 backbone, a ResBlock neck, and a YOLOv3 head. This model enabled the crop protection UAV to perform embedded real-time pest detection and autonomous spraying of pesticides. The model was tested on the Intel NCS2 hardware accelerator owing to its low weight and low power consumption. The detection system achieved an average F1-score of 92.05%.

Parico et al. [93] proposed YOLO-WEED, a weed detection system trained with 720 annotated UAV images to detect instances of weeds, based on YOLOv3 using NVIDIA GeForce GTX 1060 for green onion crops. They obtained a mAP score of 93.81% and an F1-score of 94%.

Rui et al. [94] proposed a novel comprehensive approach that combined transfer learning based on simulation data and adaptive fusion using YOLOv5 for improved detection of small objects. Their transfer learning and adaptive fusion mechanism led to a 7.1% improvement as compared to the original YOLOv5 model.

Parico et al. [95] proposed a robust real-time pear fruit counter for mobile applications using only RGB data. Various variants of YOLOv4 (YOLOv4, YOLOv4-tiny, and YOLOv4-CSP) were compared. In terms of accuracy, YOLOv4-CSP was the best model with an AP of 98%. In terms of speed and computational cost, YOLOv4-tiny showed a promising performance at a comparable rate with YOLOv4 at lower network resolutions. If considering the balance in terms of accuracy, speed, and computational cost, YOLOv4 was found to be the most suitable with AP > 96%, inference speed of 37.3 FPS, and FN Rate of 6%. Thus, YOLOv4-512 was chosen as the detection model for the pear counting system with Deep SORT.

Jintasuttisak et al. [96] exploited the effective use of YOLO-V5 in detecting date palm trees in images captured by a UAV flying above farmlands in the Northern Emirates of the United Arab Emirates (UAE). The results of using YOLO-V5 for date palm tree detection in drone imagery were compared with those obtainable with other popular CNN architectures, YOLOv3, YOLOv4, and SSD300, both quantitatively and qualitatively. The results showed that for the training data used, the YOLO-V5m (medium depth) model had the highest accuracy, resulting in a mAP of 92.34%. Further, it provided the ability to detect and localize date palm trees of varied sizes in crowded, overlapped environments and areas where the date palm tree distribution was sparse.

Tian et al. [97] proposed an anthracnose lesion detection method based on deep learning. Cycle GAN was used for data augmentation. DenseNet was then utilized to optimize feature layers of the YOLO-V3 model, which had lower resolution. The improved model exceeded Faster R-CNN with VGG16 and the original YOLO-V3 model and could realize real-time detection. The model obtained an F1-score of 81.6% and 91.7% IoU on the entire dataset.

Table 8 presents a summary of methods using YOLO. As the Table shows, it is possible to get above 90% results from most YOLO models in a variety of domains.

Table 8. YOLO Summary.

Paper	YOLO			
	Model/ Architecture	Strengths	Comments	Best Results
Chen et al. [91]	Tiny-YOLOv3	Results in excellent outcomes in regard of FPS and mAP. In addition, reduces pesticide use.	Has a high false identification in adult T. papillosa	mAP score of 95.33%
Qin et al. [92]	Ag-YOLO (v3-tiny)	Tested YOLOv3 with multiple backbones and achieved optimum results in terms of FPS and power consumption	Uses NCS2 that supports 16-bit float point values only	F1 Score of 92.05%
Parico et al. [93]	YOLO-Weed (v3)	High speed and mAP score	Limitations in detecting small objects	mAP score of 93.81% F1 score of 94%
Rui et al. [94]	YOLOv5	Improves YOLOv5 ability to detect small objects	Was not tested on agricultural applications	+7.1% of the original model
Parico et al. [95]	YOLOv4 (multiple versions)	Proves that YOLOv4-CSP has the lowest FPS with the highest mAP	Limitations in detecting small objects	AP score of 98%
Jintasuttisak et al. [96]	YOLOv5m	Compares different YOLO versions and proves that YOLOv5 with medium depth outperforms the rest even with overlapped trees.	YOLOv5x scored a higher detection average due to the increased number of layers.	mAP score of 92.34%
Tian et al. [97]	YOLOv3 (modified)	Tackles the lack of data by generating new images using CycleGAN	The model is weak without the images generated using CycleGAN	F1-score of 81.6% and IoU score of 91.7%

#### 5.2.5. Single Shot Detector (SSD)

The Single Shot Detector (SSD) is a one-stage object detection network that can detect objects in one feed-forward pass with low-resolution input images [98]. The model consists of 3 different modules. The first is a feature extraction module. This module is made up of a truncated base CNN model that is followed by convolutional layers used for the extraction of features at various scales. The second module is the object detection module which takes in feature maps and runs a set of default bounding boxes on their cells. The result is a defined number of box predictions, all of which have a shape offset and a class confidence score associated with them. The last module is the non-maximal suppression module which chooses the best predictions out of the set presented by the detection module using a specific value of IoU and confidence score as a threshold. Lately, SSDs have made an appearance in precision agriculture for their ability to perform fast inference and work with low-resolution input images. These two features of SSDs make them desirable in real-time precision agriculture applications.

Veeranampalayam Sivakumar et al. [99] proposed using a single shot detector to detect mid-to-late season weeds in soybean fields for weed-spread suppression. The authors used a feature extractor from the Inception V2 network and a stack of four extra convolutional layers to extract features at varying scales. The output of this feature extraction module were six feature maps that were then fed into the SSD's detection module. A set of bounding boxes with five different aspect ratios and six different scales were used on all locations in all six feature maps, resulting in several

box-bounded detection predictions, each prediction with its own shape offset and class confidence score. An RMS prop optimizer was used. After training the model over 25,000 epochs, the model achieved a precision of 66%, a recall of 68%, an F1-score of 67%, a mean IoU of 84%, and an inference time of 21s over 1152 x1152 image test data.

Ridho and Irwan [100] proposed a Strawberry picking robot that could detect strawberries of different health states in real-time. The robot ran an SSD-MobileNet architecture on a single board computer (SBC) to perform real-time inference. The network used a feature extraction module built with a MobileNet backbone. The choice of MobileNet was prompted by computational power and time restrictions associated with running a real-time inference model on a low-computational power single-board computer. Using transfer learning, the SSD-MobileNet V1 model was previously trained on 91 classes from the COCO dataset. The model was then re-trained on two new datasets containing a total of 250 training images of strawberries in good and bad condition. The result of the training returned an accuracy of 90% in detecting good and bad strawberries on image input extracted from a real-time-streamed video. Table 9 presents a summary of SSD methods.

Paper	SSD			
	Model/ Architecture	Strengths	Comments	Best Results
Veeranampalayam Sivakumar et al. [99]	SSD with a feature extraction module made of an Inception v2 network and 4 convolutional layers	Model is scale and translation invariant	Low optimal confidence threshold value of 0.1 Failure to detect weeds at the borders of images	Precision: 0.66, Recall: 0.68, F1 score: 0.67, Mean IoU: 0.84, Inference Time: 0.21s
Ridho and Irwan. [100]	SSD with MobileNet as a base for the feature extraction module	Fast detection and image processing	Detection was not performed on a UAV Model did not yield the best accuracy in the paper	Accuracy: 90%

5.2.6. Region-Based Convolutional Neural Networks

The Region-Based Convolutional Neural Network (R-CNN) is a two-stage object detection system that extracts many region proposals from input images, uses a CNN to perform forward propagation on each region proposal to extract its features, and then uses these features to predict the class and bounding box of this region proposal.

Sivakumar et al. [99] proposed an approach where object detection-based CNN models were trained and evaluated using low-altitude UAV images to detect weeds in mid and late seasons in soybean fields. Faster RCNN and SSD were both evaluated and compared in terms of weed detection performance. When Faster RCNN was configured with 200 box proposals, its weed detection performance was like the SSD model. The Faster RCNN model with 200 box proposals returned a precision of 0.65, a recall of 0.68, an F1-score of 0.66 and IoU of 0.85. On the other hand, the SSD model returned 0.66, 0.68, 0.67 and 0.84, for precision, recall, F1-score, and IoU, respectively. The performance of a patch-based CNN model was also evaluated and compared to the previous models. Faster RCNN model performed better than the patch-based CNN model. In conclusion Faster RCNN was found to be the best model in terms of weed detection performance and inference time among the different models compared in this study.

Ammar et al. [101] proposed an original deep learning framework for the automated counting and geolocation of palm trees from aerial images. They applied several recent convolutional neural network models (Faster R-CNN, YOLOv3, YOLOv4, and EfficientDet) to detect palm trees and other trees and conducted a complete comparative evaluation in terms of average precision and inference speed. YOLOv4 and EfficientDet-D5 yielded the best trade-off between accuracy and speed (up to 99% mAP and 7.4 FPS).

Su et al. [102] used the Mask-RCNN model for identifying Fusarium head blight disease in wheat spikes and its degree of severity. To perform this task, two Mask-RCNNs performed instance

segmentation on the input images, one of which segments individual spikes in the images, and the other segments diseased areas of spikes. Thereafter, the severity of infection of spikes was evaluated by calculating the ratio of infected spike pixels in the images to the total number of spike pixels. The backbone of this model, for feature map extraction was composed of a combination of a ResNet-101 model and an FPN model. The model returned a prediction accuracy of 77.19% after comparing the results to a set of manually labelled images.

Yang et al. [103] used an FCN-AlexNet model to perform real-time crop classification using edge computing. The authors collected 224 images using a UAV during the growing period of rice and corn. The quantitative analysis showed that the SegNet model slightly outperformed FCN-AlexNet by 1% in the overall recall rate of object classification.

Menshchikov et al. [104] proposed an approach for fast and accurate detection of hogweed. The approach includes a UAV with an embedded system on board running various Fully Convolutional Neural Networks (FCNN). They proposed an optimal architecture of FCNN for the embedded system relying on the trade-off between the detection quality and frame rate. In their pilot study, they determined that different architectures could successfully solve the semantic segmentation task for the aerial hogweed detection of two classes. The SegNet model achieved the best ROC AUC with 96.9%. This model could detect hogweed, which was not initially labeled. The Modified U-Net architecture was characterized by a high frame rate (up to 0.7 FPS) and a reasonable recognition quality (ROC AUC > 0.938). Along with the low power consumption, the U-Net architecture demonstrated its applicability for real-time scenarios and running on edge-computing devices. One of the U-Net modifications could achieve 0.46 FPS on the NVIDIA Jetson Nano platform with the ROC AUC of 0.958.

Bah et al. [77] proposed a model that combined CNN and the Hough transform to detect crop rows in images taken by a UAV. The model called CRowNet was a combination of SegNet (S-SegNet) and a CNN Hough transform (HoughCNet). The model achieved an accuracy of 93.58% and IoU of 70%, respectively.

Hosseiny et al. [10] proposed a model with the framework’s core based on a faster Regional CNN (R-CNN) model with a backbone of ResNet-101 for object detection. The proposed framework’s primary idea was to generate unlimited simulated training data from an input image automatically. The authors proposed a fully unsupervised model for plant detection in UAV-acquired pictures of agricultural fields. Two datasets were used with 442 and 328 field patches, respectively. The precision, recall, and F1-score were 0.868, 0.849, and 0.855, respectively. Table 10 shows a summary of papers using two stage detectors.

Table 10. Two-Stage Detectors Summary.

Paper	DT & RF			
	Model/ Architecture	Strengths	Comments	Best Results
Sivakumar et al. [99]	FRCNN	the optimal confidence threshold of the SSD model was found to be much lower than that of the Faster RCNN model.	Inference time of SSD is better than that of FRCNN, but it can be improved with a cost in performance.	66% F1-score and 85% IoU.
Ammar et al. [101]	FRCNN	Large advantage in terms of speed	Very weak in detecting trees. Outperformed by Efficient-Det D5 and YOLOv3	87.13% and 49.41% IoU for Palm and Trees respectively
Su et al. [102]	Mask RCNN	Superior in comparison to CNN.	Inference time was not taken into consideration	98.81% accuracy
Yang et al. [103]	FCN-AlexNet	Provides good comparison between SegNet and AlexNet	Outperformed by SegNet	88.48% Recall Rate

Paper	DT & RF			
	Model/ Architecture	Strengths	Comments	Best Results
Menshchikov et al. [104]	FCNN	Proposed method is applicable in real world scenario and the use of RGB cameras are cheaper than multi-spectral cameras	Complex algorithms compared to the multi-spectral approach	ROC AUC in segmentation: 0.96
Hosseiny et al. [10]	A model with the core of the framework based on the faster Regional CNN (R-CNN) with a backbone of ResNet-101 for object detection.	Results are good for an unsupervised method	Tested only on single object detection and automatic crop row estimation can fail due to dense plant distribution	Precision: 0.868 Recall: 0.849 F1: 0.855

5.2.7. Autoencoders

Weyner et al. [105] addressed the problem of automated, instance-level plant monitoring in agricultural fields and breeding plots. They proposed a vision-based approach to perform a joint instance segmentation of crop plants and leaves in breeding plots. They developed a CNN-based encoder-decoder network with lateral skip connections that follows a two-branch architecture with two task-specific decoders to determine the position of specific plant key points and group pixels to detect individual leaf and plant instances. Finally, they did pixel-wise instance segmentation of each crop and its associated leaves based on orthorectified RGB images captured by UAVs. Their method outperformed state-of-the-art instance segmentation approaches such as Mask-RCNN on this task. They achieved the highest score of 0.94 for AP50 at intermediate growth stages compared to 0.71 by Mask R-CNN with respect to the instance segmentation of sugar beet plants.

Lottes et al. [106] presented a novel approach for joint stem detection and crop-weed segmentation using a Fully Convolutional Network (FCN) integrating sequential information. Their proposed architecture enables the sharing of feature computations in the encoder while using two distinct task-specific decoder networks for stem detection and pixel-wise semantic segmentation of the input images. All their experiments were conducted using different generations of the BoniRob platform. BoniRob was built by BOSCH DeepField Robotics as a multi-purpose field robot for research and development applications in precision agriculture such as weed control, plant phenotyping, and soil monitoring. The system achieved the best mAP scores of 85.4%, 66.9%, 42.9%, and 50.1% for Bonn, Stuttgart, Ancona, and Eschikon datasets, respectively for stem detection and 69.7%, 58.9%, 52.9% and 44.2% mAP scores for Bonn, Stuttgart, Ancona, and Eschikon datasets, respectively for segmentation.

Su et al. [107] proposed a Deep Neural Network (DNN) that exploits the geometric location of ryegrass for the real-time segmentation of inter-row ryegrass weeds in a wheat field. Their proposed method introduced two subnets in a conventional encoder-decoder style DNN to improve segmentation accuracy. The two subnets treat inter-row and intra-row pixels differently and provide corrections to preliminary segmentation results of the conventional encoder-decoder DNN. A dataset captured in a wheat farm by an agricultural robot at different time instances was used to evaluate the segmentation performance, and the proposed method performed the best among various popular semantic segmentation algorithms (Bonnet, SegNet, PSPNet, DeepLabV3, U-Net). The proposed method ran at 48.95 FPS with a consumer-level graphics processing unit and thus is real-time deployable at camera frame rate. Their proposed model achieved the best mean accuracy and IoU scores of 96.22% and 64.21%, respectively.

Table 11. Autoencoder Summary.

Paper	Autoencoder			
	Model/ Architecture	Strengths	Comments	Best Results
[105]	CNN-Autoencoder	Performed joint instance segmentation of crop plants and leaves using a two-step approach of detecting individual instances of plants and leaves followed by pixel-wise segmentation of the identified instances.	Low segmentation precision for smaller plants. (Outperformed by Mask R-CNN)	0.94 for AP50
[106]	FCN-Autoencoder	Performed joint stem detection and crop-weed segmentation using an autoencoder with two task-specific decoders, one for stem detection and the other for pixel-wise semantic segmentation.	Did not achieve best mean recall across all tested datasets. + false detections of stems in soil regions	Achieved mAP scores of 85.4%, 66.9%,42.9%, and 50.1% for Bonn, Stuttgart, Ancona, and Eschikon datasets, respectively for stem detection and 69.7%, 58.9%, 52.9% and 44.2% mAP scores for Bonn, Stuttgart, Ancona, and Eschikon datasets, respectively for segmentation.
[107]	Autoencoder	Utilized two position-aware encoder-decoder subnets in their DNN architecture to perform segmentation of inter-row and intra-row Rygrass with higher segmentation accuracy.	Low pixel-wise semantic segmentation accuracy for early-stage wheat.	mean accuracy and IoU scores of 96.22% and 64.21%, respectively.

5.2.8. Transformers

Vaswani et al. [108] proposed the transformer architecture based on the attention mechanism. A transformer is a sequence transduction model initially designed to tackle natural language processing (NLP) problems. Using transformers for computer vision tasks was limited initially due to the high computational cost of training. To address this issue, Dosovitskiy et al. [109] proposed the Vision Transformer (ViT) that requires fewer resources while out-performing convolutional networks (CNN). Other notable contributions include utilizing Detection Transformers (DETR) targeting the same problem. [110].

Thai et al. [111] used ViTs for the early detection of infected cassava leaves and the classification of their diseases. Initially, they used the ImageNet pre-trained ViT model published by the Google Research Team [112]. The model was then tuned using Cassava Leaf Disease Dataset [113]. Later, the model was quantized to reduce its size and accelerate the inference step (FPS) before deploying it on a Raspberry Pi 4 Model B. Their model achieved a 90.3% F1-score in comparison to the best CNN score of 89.2% achieved by the Resnet50 model. Further, they proposed a smart solution powered by the Internet of Things (IoT) that can be used in the agriculture industry for real-time detection of leaf diseases. The system consists of a drone that captures the leaf images, including the exact position of the spot in the field. The ViT model installed on the Drones Pi classifies the images and clusters the infected leaves. The results are then combined with the spot’s position and sent to a server via a 4G network to create a survey map of the field. Farmers and rescue agencies can obtain the map on their mobile phones and prevent the loss of crops beforehand.

Reedha et al. [23] used two different models of ViT for plant classification of UAV images. Images were collected using a drone mounted with a high-resolution camera and deployed in a crop

field of beet, parsley, and spinach located in France. The camera captured RGB orthorectified images at regular intervals in the field. The data was manually labelled into five classes: Weeds, Beet, Parsley, Spinach, and Off-type green leaves. They also employed data augmentation to help improve the robustness of the model and the generalization capabilities of the training dataset. Later, they used ViT-B32 and ViTB16 models. They also tested the training data on EfficientNet and ResNet CNN architectures for comparison purposes. The results showed that ViT models outperformed the CNN models as an F1-score of 99.4% and 99.2% were obtained from ViT-B16 and ViT-B32, respectively. In comparison, CNN models achieved slightly lower scores of 98.7% for EfficientNet B0, 98.9% for B1, and a close 99.2% using ResNet50. The authors pointed out that although all techniques obtained high accuracy and F1-scores, the classification of crops and weed images using ViTs yielded the best prediction performance. However, the inefficiency of ViT as compared with CNNs is another consideration if the model is to be deployed for real-time processing on a UAV.

Karila et al. [114] used ViT models to estimate grass sward (i.e., short grass) quality and quantity in a field. The datasets were captured in the spring “primary growth phase,” and the same dataset was captured again in the summer “regrowth phase” using a quadcopter drone equipped with two cameras. The first captured RGB images, while the second captured Fabry Perot (FPI) images. The results showed that ViT RGB models performed the best on different datasets. Similarly, VGG CNN models provided equally satisfactory results in most cases.

Dersch et al. [115] used a detection transformer (DETR) to detect single trees in high-resolution RGB true orthophotos (TDOPs) and compared it to a YOLOv4 single-stage detector. The multispectral images were collected by a ten channels camera system with a horizontal field of view. Later, the images were post-processed using structure-from-motion (SFM) software. The data was later manually labelled with a split of 80% training and 20% for validation. DETR outperformed YOLOv4 in mixed and deciduous plots with a 20% difference in F1-score in mixed plots and 4% in the latter plots; 86% to 65% and 71% to 67%, respectively. Across all three test plots, both methods had problems with over-segmentation. Furthermore, DETR failed to detect smaller trees far worse than YOLOv4 in multiple cases. The authors justified these poor results by the fact that DETR uses lower resolution feature maps than that of YOLOv4.

Chen et al. [116] proposed a new efficient deep learning model called the Density Transformer (DENT) for automatic tree counting from aerial images. The model’s architecture contains four stages: a Multi-Receptive Field CNN (Multi-RF CNN) to compute a feature map over the input images, followed by a standard transformer encoder, and a Density Map Generator (DMG) to predict the density distribution over the input images. They also introduced a benchmark dataset that contains aerial images for tree counting called Yosemite Tree dataset and released it to the public [116]. The model outperformed most state-of-the-art methods with a MAE of 10.7 and a RMSE of 13.7 in comparison to 17.3 and 22.6, respectively, using YOLOv3. It is worth mentioning that the CANNet model [117] achieved the closest values with 10.8 and 13.8, respectively, and achieved a better MAE score in one of four regions than the DENT models.

Finally, Zhang et al. [118] developed a spectral-spatial attention-based transformer (SSVT) to estimate crop nitrogen status from UAV imagery. The model is an improved version of the standard Vision Transformer (ViT) that can extract the spatial information of images. The newly proposed model can predict the spectral information which contains most of the features in agricultural applications. The model also tackles the computational complexity of large images that ViT suffers from by adopting a self-supervised learning (SSL) technology to allow models to train with unlabeled data. The results showed that the model with 96.2% accuracy outperformed the ViT model with 94.4% accuracy. However, this model required 4 million additional parameters to those required for a ViT model. Table 12 presents a summary of methods using transformers.

Table 12. Transformers Summary.

Paper	Transformers			
	Model/ Architecture	Strengths	Comments	Best Results
Reedha et al. [23]	ViT	The classification of crops and weed images using ViTs yields the best prediction performance.	Slightly overperformers existing CNN models	an F1-score of 99.4% and 99.2% were obtained from ViT-B16 and ViT-B32, respectively.
Karila et al. [114]	ViT	The ViT RGB models performed the best on several types of datasets.	VGG CNN models provided equally reliable results in most cases.	Multiple results shown on several types of datasets
Dersch et al. [115]	DETR	DETR clearly outperformed YOLOv4 in mixed and deciduous plots	DETR failed to detect smaller trees far worse than YOLOv4 in multiple cases.	An F1-Score of 86% and 71% in mixed and deciduous plots respectively
Chen et al. [116]	DENT	The model outperformed most of the state-of-the-art methods	CANNet achieved better results	a Mean Absolute Error (MAE) of 10.7 and a Root Mean Squared Error (RMSE) of 13.7
Coletta et al [119]	Active Learning	The model can classify unknown data.	Did not test the performance of different classification models other than their own.	An accuracy of 98% and a Recall of 97%

5.2.9. Active Learning

Coletta et al. [119] used a semi-supervised classification algorithm that can aggregate information from clusters with those provided by a supervised algorithm like SVM to discover new classes in an active learning manner. According to the authors, such an ability is largely convenient for inconsistent agricultural environments. The data was collected through a SenseFly eBee equipped with an RGB camera. The model consisted of two blocks; a classification block (ClaB) representing an area of 0.16m<sup>2</sup> to be classified and a contextual block (ConB) providing supplementary context information. Both blocks formed a concentric pair that generates feature vectors to be classified. These vectors were manually labeled as belonging to one of three classes. Then a semi-supervised classifier was used to quantify the uncertainty of classification, and a density measure evaluated the importance of a classified feature vector. If the instances resulted in high uncertain labels, they were denoted as novelties to be learned, which was labelled later by an Entropy and Density-based Selection (EDS) domain expert and incorporated into the training set. The results showed that the all-class accuracy and recall improved iteratively.

5.2.10. Light Detection and Ranging Algorithm (LiDAR)

for individual tree detection from UAV imagery. The model used an existing LIDAR algorithm to generate RGB trees that could be used for training as a starting point. The model was then retained using a small number of manual labels to correct errors from the unsupervised detection. Then a pre-trained ResNet50 backbone was used to classify the images. The model was tested on the NEON public dataset and achieved the best performance among existing LIDAR-based models (+2%) in comparison to that achieved by Silva et al. [120].

### 5.2.10. Semi-supervised Convolutional Neural Networks

Bosilj et al. [121] used the fundamental SegNet architecture to perform pixel-level classification and segmentation of three classes of soil. The input comprises RGB and near-infrared (NIR) images. The authors used a median frequency weighting to avoid unbalanced labelling as soil pixels are dominant in any given field with respect to crops or weeds. The input data was taken in the form of RGB and NIR channels directly because NDVI preprocessing typically results in minimal differences. The model was trained on three different datasets of sugar beets, carrots, and onions (SB16, CA17, ON17) in which there were fully labelled examples in one, and partially labelled examples in the other and on both pixel-level and object-level training. Object-based detection performed better than pixel-based detection precision-wise. However, pixel-based detection performed better in recall. It is worth noting that the partially labelled ON17 dataset with SB16 weights outperformed the fully labelled dataset. The partially labelled CO17 dataset performed significantly worse than the fully labelled dataset, with a difference of almost 20% on weeds and 5% on crops.

## 6. Discussion and Future Work

### 6.1. Machine Learning Techniques

In general, SVMs did not work well in comparison to deep learning approaches. References [47], [48], [56], [57], [122] used SVMs to classify crops/weeds in agricultural fields. Most of the results showed low accuracies. This can be because SVMs underperform when there is no clear margin of difference between the different classes, which is usually the case in agricultural imagery even with the images being preprocessed. In addition, SVMs are more likely to fail when classes are noisy and overlapping. KNN suffered from similar limitations. References [48], [58], [59] showed that KNNs performed slightly better than SVMs. Nonetheless, they were still sensitive to noisy data. Random Forests (RF) were used in [49], [56] and performed better than SVM and KNN in this limited context. However, RF require higher computational cost as the algorithm involves multiple decision trees within, which makes it challenging to implement on a UAV for real-time predictions. Decision trees and RF also have the problem of overfitting. It is worth mentioning that Coletta et al. [119] used active learning to discover new classes through an SVM technique on semi-labelled data, and the results were promising and reliable. The Light Detection and Learning algorithm was used by Weinstein [123] with a ResNet backbone and achieved decent results as well.

CNNs represent a good candidate for solving image-based classification and detection problems in precision agriculture. U-Net models performed well with fewer training samples and provided better performance for segmentation tasks [124]. References [24], [80]–[84], [89] showed that U-Net outperformed other CNN models. In addition, Arun et al. [24] showed that U-Net can be further optimized without compromising on performance. Other architectures included ResNet, ShuffleNet, ShuffleNetV2, and MobileNet, which all require larger computational costs and are less suitable for real-time UAV applications.

Single Stage Detectors like the YOLO series, CornerNet, and CenterNet improved the detection speed while maintaining high accuracy. Tiny-YOLOv3 worked well in real-time applications due to a small number of parameters, high speed, and efficient computation. References [91]–[93], [97] showed that YOLOv3 performed the best among all YOLO models. Using Tiny-YOLOv3 resulted in a slight trade-off with accuracy, especially with detecting smaller objects. Nonetheless, the overall accuracy remained high. Two-Stage Detectors such as R-CNN, FPN, and Mask R-CNN performed better than Single-Stage detectors. References [76], [99], [100], [102], [105]–[107] showed that these models outperformed single-stage detectors in terms of recall and accuracy. However, many authors argued that region proposal modules required higher computation and run-time memory footprint, thus making detection slow even on high-end GPUs [92].

Transformers represent a viable approach in agricultural classification tasks using UAV image data. In specific, the ViT model showed promising results. References [23], [111], [114] compared ViT models with current CNN architectures and showed that both approaches achieved similar results, with ViTs enjoying a slight edge. The DETR model [115] was compared to the YOLO series of models

and both approaches achieved similar results as well. However, it was evident that DETR models fell short in detecting smaller trees and crops due to their encoder-decoder nature. DENT was used in [116] and outperformed most current methods. However, the CNN model CANNet achieved better results on the same data.

Generative Adversarial Networks (GANs) were primarily used to enhance the training process by adding to the manually labelled data in a semi-supervised manner. References [125], [126] used semi-supervised GANs (SGAN) and cGAN, respectively. In both works, GAN architectures were outperformed by CNNs for higher labelling rates.

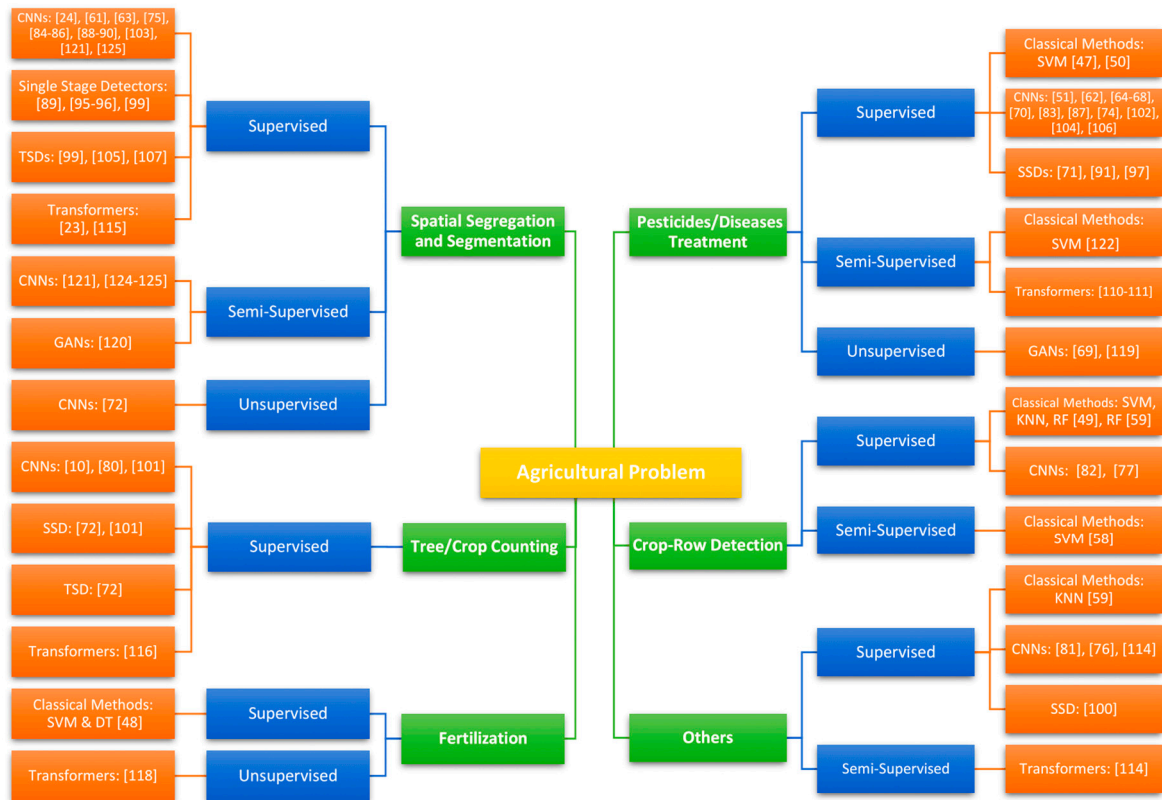
## 6.2. Best Techniques for Agricultural Problems

Table 13 shows the current best solutions for each problem and the respective type of learning architecture used. Because of the unavailability of appropriate benchmarks, it is difficult to compare the proposed approaches. However, Table XIII represents the best results achieved using specific data sets. The results generally show that machine learning and deep learning can yield reasonable results for a variety of problems. There is clearly room for improvement in most cases as the results are sometimes in the range of 80% accuracy.

**Table 13.** Best Results Achieved in Different Agricultural Problems.

<i>Problem</i>	<i>Type of learning</i>	<i>Paper</i>	<i>Model/Architecture</i>	<i>Dataset</i>	<i>Best Results</i>
Spatial Segregation and Segmentation	Supervised	Jintasuttisak et al. [96]	YOLOv5	date palm trees collected using a drone	mAP score of 92.34%
	Semi-Supervised	Fawakherji et al [118]	cGANs	of 5,400 RGB images of pears and strawberries of which 20% were labelled	An IoU score of 83.1% on mixed data including both original and synthesized.
	Unsupervised	Bah et al. [72]	ResNet18	UAV images of spinach and bean fields	AUC: 91.7%
Pesticides/ Diseases Treatment	Supervised	Zhang et al. [73]	DF-U-Net	Yangling UAV images	F1: 94.13% Accuracy: 96.93% Precision: 94.02%
	Semi-Supervised	Coletta et al [119]	Active Learning: SVM	UAV images collected from Eucalyptus plantations	An accuracy of 98% and a Recall of 97%
	Unsupervised	Khan et al. [117]	SGAN	UAV images collected from strawberry and pea fields	Accuracy ~90%
Fertilization	Supervised	Natividade et al. [56]	SVM	UAV images of vineyards and forests	Accuracy: 83%, Precision: 97%, Recall: 94%
	Unsupervised	Zhang et al. [115]	SSVT	UAV images of a wheat field	96.5% on 384x384-sized images
Crop-Row Detection	Supervised	Cesar Pereira et al. [47]	SVM	Manually collected RGB images	88.01% F1-Score
	Semi-Supervised	Pérez-Ortiz et al. [57]	SVM	UAV images collected from a sunflower plot	MAE: 12.68%
Tree/Crop Counting	Supervised	Ammar et al. [98]	FRCNN	Tree counting	87.13% IoU om Palms and 49.41% on other trees
	Semi- Supervised	Chen et al. [113]	DENT	Yosemite Tree dataset	10.7 MAE score
Others	Supervised	Aiger et al. [76]	CNN	UAV images of various types of landcover	96.3% accuracy

Figure 4 shows an overall summary of machine learning techniques using UAV image data for precision agriculture. As the Figure shows, supervised, semi-supervised and unsupervised techniques have been used for a variety of problems.



**Figure 4.** A Summary of the papers surveyed based on the agricultural problem and technique used.

### 6.3. Future Work

The current state of practice in deploying UAVs in an agricultural field typically consists of using multiple UAVs or stages. One smaller UAV is used to collect images from the field. Once the images are collected, machine learning algorithms are applied, and the results are used to program another typically larger delivery UAV that applies pesticides, fertilizer or similar in a smarter fashion. A future vision is to have autonomous agricultural UAVs that can process images on-board and take appropriate actions as necessary. However, current UAVs are resource-constrained, and their performance is limited by energy consumption, memory size, and latency. As a result, it is not practical to use high-resource algorithms to perform detection and classification tasks. Such issues can be addressed by using low-bit architectures, by compressing a dense model, by using an effective model with a small number of parameters, or by using a hardware accelerator that can be deployed on an embedded System-on-Chip (SoC) that includes graphic processing units (GPUs) or Field-programmable gate arrays (FPGAs) [127].

As the survey showed, many generic CNN backbones (e.g., ResNet18) have been used to address various agricultural problems. However, these backbones are typically pre-trained on non-agricultural data and are only finetuned on agricultural data. Availability of agricultural benchmarking data sets will help make progress towards creating pre-trained backbones for agricultural problems.

Many agricultural problems require semantic segmentation, object recognition, and instance segmentation. As Table XII shows, single stage detectors like YOLOv3 have performed well for object recognition. In addition, two stage detectors like FRCNN and specialized architectures like DENT have relatively better performance. Transformers like ViT have been used. However, one issue with transformer-based architectures is the higher computational inefficiency as compared to CNNs. Many new approaches to optimize transformers for a more efficient footprint have been proposed [128]. This work is highly relevant for autonomous UAVs that need to perform the inference onboard the UAV. In addition, entirely new architectures like Hyena [129] that claim sub-quadratic performance when compared with transformers have been proposed. Such new architectures also

carry promise with combined with the current object recognition and instance segmentation approaches.

Many surveyed UAV studies used multispectral data as opposed to using just the RGB data. Multispectral data represents a special challenge for UAVs because using higher number of channels in the input significantly increases the memory requirements which is not ideal for autonomous UAVs. Building efficient CNNs for multispectral data is a well-researched problem [130]. Transformers are also being used with multispectral data (e.g., [131]). Approaches for segmentation using multispectral data (e.g., [132], [133]) have been recently proposed as well. Efficient handling of multispectral data when using transformers represents another important research problem for the UAV image data.

Finally, a significant amount of work has been done in agriculture using satellite imagery [134]. The appearance of multimodal agricultural data including synced and UAV images (e.g., [135]) raises the interesting possibility of multimodal systems on a UAV utilizing ground, UAV and satellite data in tandem to solve the various agricultural problems more effectively.

From the agricultural problem perspective, many of the problems identified in Table I have been addressed using image data from UAVs. However, some problems in the post-harvesting stage like fruit grading, quality retention, storage environments, etc. may require additional attention.

## 7. Conclusion

In this survey, over 70 recent papers using UAV agricultural imagery to classify, detect, and segment crops and trees using machine learning algorithms and deep learning models were discussed. Deep learning models like U-Net, YOLOv3, and ViT performed the best among state-of-the-art approaches. The primary challenges include detecting small trees, interleaved crops, and the high-power consumption of complex models. Future work includes developing low power consuming and less expensive models that could be deployed on the UAVs to perform real time or on edge tasks that can provide faster and more sufficient solutions in the field of precision farming.

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

## References

1. "2022 Global Report on Food Crises:" p. 277.
2. T. Ayoub Shaikh, T. Rasool, and F. Rasheed Lone, "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming," *Computers and Electronics in Agriculture*, vol. 198, p. 107119, Jul. 2022, doi: 10.1016/j.compag.2022.107119.
3. I. Mylonas, D. Stavrakoudis, D. Katsantonis, and E. Korpetis, "Better farming practices to combat climate," *Academic Press*, pp. 1–29, 2020, doi: 10.1016/B978-0-12-819527-7.00001-7.
4. A. Wolińska, "Metagenomic Achievements in Microbial Diversity Determination in Croplands: A Review," in *Microbial Diversity in the Genomic Era*, S. Das and H. R. Dash, Eds., Academic Press, 2019, pp. 15–35. doi: 10.1016/B978-0-12-814849-5.00002-2.
5. Z. Mohamed, R. Terano, J. Sharifuddin, and G. Rezai, "Determinants of Paddy Farmer's Unsustainability Farm Practices," *Agriculture and Agricultural Science Procedia*, vol. 9, pp. 191–196, Jan. 2016, doi: 10.1016/j.aaspro.2016.02.120.
6. K. R. Krishna, *Push button agriculture: Robotics, drones, satellite-guided soil and crop management*. CRC Press, 2017.
7. ISPA, "Precision Ag Definition | International Society of Precision Agriculture." <https://www.ispag.org/about/definition> (accessed Jul. 14, 2022).
8. P. Singh *et al.*, "8 - Hyperspectral remote sensing in precision agriculture: present status, challenges, and future trends," in *Hyperspectral Remote Sensing*, P. C. Pandey, P. K. Srivastava, H. Balzter, B. Bhattacharya, and G. P. Petropoulos, Eds., in Earth Observation. Elsevier, 2020, pp. 121–146. doi: 10.1016/B978-0-08-102894-0.00009-7.
9. I. Cisternas, I. Velásquez, A. Caro, and A. Rodríguez, "Systematic literature review of implementations of precision agriculture," *Computers and Electronics in Agriculture*, vol. 176, p. 105626, Sep. 2020, doi: 10.1016/j.compag.2020.105626.
10. B. Hosseiny, H. Rastiveis, and S. Homayouni, "An Automated Framework for Plant Detection Based on Deep Simulated Learning from Drone Imagery," *Remote Sensing*, vol. 12, no. 21, Art. no. 21, Jan. 2020, doi: 10.3390/rs12213521.

11. R. Y. Aburasain, E. A. Edirisinghe, and A. Albatay, "Palm Tree Detection in Drone Images Using Deep Convolutional Neural Networks: Investigating the Effective Use of YOLO V3," in *Digital Interaction and Machine Intelligence*, C. Biele, J. Kacprzyk, J. W. Owsinski, A. Romanowski, and M. Sikorski, Eds., in *Advances in Intelligent Systems and Computing*. Cham: Springer International Publishing, 2021, pp. 21–36. doi: 10.1007/978-3-030-74728-2\_3.
12. Y. Lu and S. Young, "A survey of public datasets for computer vision tasks in precision agriculture," *Computers and Electronics in Agriculture*, vol. 178, p. 105760, Nov. 2020, doi: 10.1016/j.compag.2020.105760.
13. U. Shafi, R. Mumtaz, J. García-Nieto, S. A. Hassan, S. A. R. Zaidi, and N. Iqbal, "Precision Agriculture Techniques and Practices: From Considerations to Applications," *Sensors*, vol. 19, no. 17, Art. no. 17, Jan. 2019, doi: 10.3390/s19173796.
14. A. Kamilaris and F. X. Prenafeta-Boldú, "A review of the use of convolutional neural networks in agriculture," *J. Agric. Sci.*, vol. 156, no. 3, pp. 312–322, Apr. 2018, doi: 10.1017/S0021859618000436.
15. D.-K. K. Chengjuan Ren and Dongwon Jeong, "A Survey of Deep Learning in Agriculture: Techniques and Their Applications," *Journal of Information Processing Systems*, vol. 16, no. 5, pp. 1015–1033, Oct. 2020, doi: 10.3745/JIPS.04.0187.
16. V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. Ramkteke D., "Machine learning in agriculture domain: A state-of-art survey | Elsevier Enhanced Reader." <https://reader.elsevier.com/reader/sd/pii/S2667318521000106?token=634DCBF0EC91ABEE41DEF3AF831054DC77F10DB01949D5414FBFDEC54E71CE2F2127753BEBEF4695C760D0F00D992D14&originRegion=eu-west-1&originCreation=20220719054056> (accessed Jul. 19, 2022).
17. P. Radoglou-Grammatikis, P. Sarigiannidis, T. Lagkas, and I. Moscholios, "A compilation of UAV applications for precision agriculture," *Computer Networks*, vol. 172, p. 107148, May 2020, doi: 10.1016/j.comnet.2020.107148.
18. M. F. Aslan, A. Durdu, K. Sabanci, E. Ropelewska, and S. S. Gültekin, "A Comprehensive Survey of the Recent Studies with UAV for Precision Agriculture in Open Fields and Greenhouses," *Applied Sciences*, vol. 12, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/app12031047.
19. "FAO - News Article: New standards to curb the global spread of plant pests and diseases." <https://www.fao.org/news/story/en/item/1187738/icode/> (accessed Jul. 21, 2022).
20. A. Khakimov, I. Salakhutdinov, A. Omonlikov, and S. Utagnov, "Traditional and current-prospective methods of agricultural plant diseases detection: A review." [https://www.researchgate.net/publication/357714060\\_Traditional\\_and\\_current-prospective\\_methods\\_of\\_agricultural\\_plant\\_diseases\\_detection\\_A\\_review](https://www.researchgate.net/publication/357714060_Traditional_and_current-prospective_methods_of_agricultural_plant_diseases_detection_A_review) (accessed Jul. 21, 2022).
21. Y. Fang and R. P. Ramasamy, "Current and Prospective Methods for Plant Disease Detection," *Biosensors (Basel)*, vol. 5, no. 3, pp. 537–561, Aug. 2015, doi: 10.3390/bios5030537.
22. "Ecological Understanding of Insects in Organic Farming Systems: How Insects Damage Plants | eOrganic." <https://eorganic.org/node/3151> (accessed Jul. 21, 2022).
23. R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, "Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images," *Remote Sensing*, vol. 14, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/rs14030592.
24. R. A. Arun, S. Umamaheswari, and A. V. Jain, "Reduced U-Net Architecture for Classifying Crop and Weed using Pixel-wise Segmentation," in *2020 IEEE International Conference for Innovation in Technology (INOCON)*, Nov. 2020, pp. 1–6. doi: 10.1109/INOCON50539.2020.9298209.
25. A. Sharma, "Worldwide pesticide usage and its impacts on ecosystem," p. 16.
26. M. Rai and A. Ingle, "Role of nanotechnology in agriculture with special reference to management of insect pests," *Appl Microbiol Biotechnol*, vol. 94, no. 2, pp. 287–293, Apr. 2012, doi: 10.1007/s00253-012-3969-4.
27. "Pest control efficiency in Agriculture - Futurcrop." <https://futurcrop.com/en/blog/post/pest-control-efficiency-in-agriculture> (accessed Jul. 28, 2022).
28. "Urban vegetation classification: Benefits of multitemporal RapidEye satellite data | Elsevier Enhanced Reader." <https://reader.elsevier.com/reader/sd/pii/S0034425713001429?token=8C76FFADC2BF1E1A573FBD896D61E55BF50B5D27C5866C1CBB6B247EC2328C2170337BF8AF4B02F0D77FBB9520D68A30&originRegion=us-east-1&originCreation=20220722111257> (accessed Jul. 22, 2022).
29. "Remote sensing of urban environments: Special issue | Elsevier Enhanced Reader." <https://reader.elsevier.com/reader/sd/pii/S0034425711002781?token=418370169508BB5AAFF98E8B53A3FB30990D7886B72D5D22A007F023EBF3BF79DCFC674C589CA16CD7EBA979A63F179&originRegion=us-east-1&originCreation=20220722111729> (accessed Jul. 22, 2022).
30. IOWAAGLITERACY, "Why do they do that? – Estimating Yields," *Iowa Agriculture Literacy*, Sep. 18, 2019. <https://iowaagliteracy.wordpress.com/2019/09/18/why-do-they-do-that-estimating-yields/> (accessed Jul. 21, 2022).
31. T. Horie, M. Yajima, and H. Nakagawa, "Yield forecasting," *Agricultural Systems*, vol. 40, no. 1, pp. 211–236, Jan. 1992, doi: 10.1016/0308-521X(92)90022-G.

32. "Crop Yield," *Investopedia*. <https://www.investopedia.com/terms/c/crop-yield.asp> (accessed Jul. 21, 2022).
33. M. Altalak, M. Ammad uddin, A. Alajmi, and A. Rizg, "Smart Agriculture Applications Using Deep Learning Technologies: A Survey," *Applied Sciences*, vol. 12, no. 12, Art. no. 12, Jan. 2022, doi: 10.3390/app12125919.
34. R. K. Prange, "Pre-harvest, harvest and post-harvest strategies for organic production of fruits and vegetables." [https://www.researchgate.net/publication/284871289\\_Pre-harvest\\_harvest\\_and\\_post-harvest\\_strategies\\_for\\_organic\\_production\\_of\\_fruits\\_and\\_vegetables](https://www.researchgate.net/publication/284871289_Pre-harvest_harvest_and_post-harvest_strategies_for_organic_production_of_fruits_and_vegetables) (accessed Jul. 25, 2022).
35. "(10) (PDF) Tomato Fruit Yields and Quality under Water Deficit and Salinity." [https://www.researchgate.net/publication/279642090\\_Tomato\\_Fruit\\_Yields\\_and\\_Quality\\_under\\_Water\\_Deficit\\_and\\_Salinity](https://www.researchgate.net/publication/279642090_Tomato_Fruit_Yields_and_Quality_under_Water_Deficit_and_Salinity) (accessed Jul. 30, 2022).
36. E. Atay, B. Hucbourg, A. Drevet, and P.-E. Lauri, "Investigating effects of over-irrigation and deficit irrigation on yield and fruit quality in pink ladytm 'rosy glow' apple," *Acta Scientiarum Polonorum: Hortorum Cultus*, vol. 16, no. 4, pp. 45–51, 2017, doi: 10.24326/asphc.2017.4.5.
37. X. Li, Y. Ba, M. Zhang, M. Nong, C. Yang, and S. Zhang, "Sugarcane Nitrogen Concentration and Irrigation Level Prediction Based on UAV Multispectral Imagery," *Sensors*, vol. 22, no. 7, Art. no. 7, Jan. 2022, doi: 10.3390/s22072711.
38. S. Aden, J. Bialas, Z. Champion, E. Levin, and J. McCarty, *Low cost infrared and near infrared sensors for UAVS*, vol. XL-1. 2014. doi: 10.5194/isprsarchives-XL-1-1-2014.
39. I. K. Arah, H. Amaglo, E. K. Kumah, and H. Ofori, "Preharvest and Postharvest Factors Affecting the Quality and Shelf Life of Harvested Tomatoes: A Mini Review," *International Journal of Agronomy*, vol. 2015, pp. 1–6, 2015, doi: 10.1155/2015/478041.
40. "What methods can improve crop performance? | Royal Society." <https://royalsociety.org/topics-policy/projects/gm-plants/what-methods-other-than-genetic-improvement-can-improve-crop-performance/> (accessed Jul. 29, 2022).
41. T. Takamatsu *et al.*, "Over 1000 nm Near-Infrared Multispectral Imaging System for Laparoscopic In Vivo Imaging," *Sensors (Basel)*, vol. 21, no. 8, p. 2649, Apr. 2021, doi: 10.3390/s21082649.
42. A. Sellami and S. Tabbone, "Deep neural networks-based relevant latent representation learning for hyperspectral image classification," *Pattern Recognition*, vol. 121, p. 108224, Jan. 2022, doi: 10.1016/j.patcog.2021.108224.
43. "Multispectral Image - an overview | ScienceDirect Topics." <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/multispectral-image> (accessed Dec. 13, 2022).
44. S. Seager, E. L. Turner, J. Schafer, and E. B. Ford, "Vegetation's Red Edge: A Possible Spectroscopic Biosignature of Extraterrestrial Plants," *Astrobiology*, vol. 5, no. 3, pp. 372–390, Jun. 2005, doi: 10.1089/ast.2005.5.372.
45. "Land Management Information Center - Minnesota Land Ownership," *MN IT Services*. <https://www.mngeo.state.mn.us/chouse/airphoto/cir.html> (accessed Dec. 13, 2022).
46. GISGeography, "What is NDVI (Normalized Difference Vegetation Index)?," *GIS Geography*, May 09, 2017. <https://gisgeography.com/ndvi-normalized-difference-vegetation-index/> (accessed Dec. 13, 2022).
47. A. Tendolkar, A. Choraria, M. M. Manohara Pai, S. Girisha, G. Dsouza, and K. S. Adithya, "Modified crop health monitoring and pesticide spraying system using NDVI and Semantic Segmentation: An AGROCOPTER based approach," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, Aug. 2021, pp. 1–5. doi: 10.1109/ICAS49788.2021.9551116.
48. P. C. P. Júnior, A. Monteiro, R. D. L. Ribeiro, A. C. Sobieranski, and A. V. Wangenheim, "Comparison of Supervised Classifiers and Image Features for Crop Rows Segmentation on Aerial Images." <https://www.tandfonline.com/doi/epub/10.1080/08839514.2020.1720131?needAccess=true> (accessed Jun. 15, 2022).
49. P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss, "UAV-based crop and weed classification for smart farming," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 3024–3031. doi: 10.1109/ICRA.2017.7989347.
50. V. Lakshmanan, M. Görner, and R. Gillard, *Practical Machine Learning for Computer Vision*. O'Reilly Media, Inc., 2021.
51. A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "Deep learning techniques to classify agricultural crops through UAV imagery: a review," *Neural Comput & Applic*, vol. 34, no. 12, pp. 9511–9536, Jun. 2022, doi: 10.1007/s00521-022-07104-9.
52. S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digital Signal Processing*, vol. 126, p. 103514, Jun. 2022, doi: 10.1016/j.dsp.2022.103514.
53. D. Tian, Y. Han, B. Wang, T. Guan, H. Gu, and W. Wei, "Review of object instance segmentation based on deep learning," *JEL*, vol. 31, no. 4, p. 041205, Dec. 2021, doi: 10.1117/1.JEI.31.4.041205.

54. J. Zhang *et al.*, "Segmenting Purple Rapeseed Leaves in the Field from UAV RGB Imagery Using Deep Learning as an Auxiliary Means for Nitrogen Stress Detection," *Remote Sensing*, vol. 12, no. 9, Art. no. 9, Jan. 2020, doi: 10.3390/rs12091403.
55. T. N. Liliane and M. S. Charles, *Factors Affecting Yield of Crops*. IntechOpen, 2020. doi: 10.5772/intechopen.90672.
56. J. Natividade, J. Prado, and L. Marques, "Low-cost multi-spectral vegetation classification using an Unmanned Aerial Vehicle," in *2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Apr. 2017, pp. 336–342. doi: 10.1109/ICARSC.2017.7964097.
57. M. Pérez-Ortiz, J. M. Peña, P. A. Gutiérrez, J. Torres-Sánchez, C. Hervás-Martínez, and F. López-Granados, "A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method," *Applied Soft Computing*, vol. 37, pp. 533–544, Dec. 2015, doi: 10.1016/j.asoc.2015.08.027.
58. E. C. Rodríguez-Garlito and A. Paz-Gallardo, "Efficiently Mapping Large Areas of Olive Trees Using Drones in Extremadura, Spain," *IEEE Journal on Miniaturization for Air and Space Systems*, vol. 2, no. 3, pp. 148–156, Sep. 2021, doi: 10.1109/JMASS.2021.3067102.
59. B. M. Rocha, G. da Silva Vieira, A. U. Fonseca, H. Pedrini, N. M. de Sousa, and F. Soares, "Evaluation and Detection of Gaps in Curved Sugarcane Planting Lines in Aerial Images," in *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Aug. 2020, pp. 1–4. doi: 10.1109/CCECE47787.2020.9255701.
60. M. Crimaldi, V. Cristiano, A. De Vivo, M. Iernia, P. Ivanov, and F. Sarghini, "Neural Network Algorithms for Real Time Plant Diseases Detection Using UAVs," in *Innovative Biosystems Engineering for Sustainable Agriculture, Forestry and Food Production*, A. Coppola, G. C. Di Renzo, G. Altieri, and P. D'Antonio, Eds., in *Lecture Notes in Civil Engineering*. Cham: Springer International Publishing, 2020, pp. 827–835. doi: 10.1007/978-3-030-39299-4\_89.
61. A. Milioto, P. Lottes, and C. Stachniss, "Real-Time Blob-Wise Sugar Beets Vs Weeds Classification for Monitoring Fields Using Convolutional Neural Networks," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W3, pp. 41–48, 2017, doi: <https://doi.org/10.5194/isprs-annals-IV-2-W3-41-2017>.
62. M. D. Bah, E. Dericquebourg, A. Hafiane, and R. Canals, "Deep Learning Based Classification System for Identifying Weeds Using High-Resolution UAV Imagery," in *Intelligent Computing*, K. Arai, S. Kapoor, and R. Bhatia, Eds., in *Advances in Intelligent Systems and Computing*, vol. 857. Cham: Springer International Publishing, 2019, pp. 176–187. doi: 10.1007/978-3-030-01177-2\_13.
63. S. R. G. Reddy, G. P. S. Varma, and R. L. Davuluri, "Optimized convolutional neural network model for plant species identification from leaf images using computer vision," *Int J Speech Technol*, Apr. 2021, doi: 10.1007/s10772-021-09843-x.
64. A. Sembiring, Y. Away, F. Arnia, and R. Muharar, "Development of Concise Convolutional Neural Network for Tomato Plant Disease Classification Based on Leaf Images," *J. Phys.: Conf. Ser.*, vol. 1845, no. 1, p. 012009, Mar. 2021, doi: 10.1088/1742-6596/1845/1/012009.
65. Geetharamani G. and Arun Pandian J, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network | Elsevier Enhanced Reader." <https://reader.elsevier.com/reader/sd/pii/S0045790619300023?token=2C657881C2D2EDBB85C23044E8BFC A69B649314432597FA9D45A48392FD1B10F49E30D04D95706CA328CD1A114A7E17E&originRegion=eu-west-1&originCreation=20220615171346> (accessed Jun. 15, 2022).
66. K. R., H. M., S. Anand, P. Mathikshara, A. Johnson, and M. R., "Attention embedded residual CNN for disease detection in tomato leaves," *Applied Soft Computing*, vol. 86, p. 105933, Jan. 2020, doi: 10.1016/j.asoc.2019.105933.
67. L. Nanni, A. Manfè, G. Maguolo, A. Lumini, and S. Brahmam, "High performing ensemble of convolutional neural networks for insect pest image detection," *Ecological Informatics*, vol. 67, p. 101515, Mar. 2022, doi: 10.1016/j.ecoinf.2021.101515.
68. Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model | Elsevier Enhanced Reader." <https://reader.elsevier.com/reader/sd/pii/S1574954120301321?token=6A5F2FC8AE4A263722309CBE19B19 5E31EB4D614A298D3B3DAD91C75A971E4A88E7545F453E28200C3D2E11D49D8C189&originRegion=us-east-1&originCreation=20220616151758> (accessed Jun. 16, 2022).
69. A. Prasad, N. Mehta, M. Horak, and W. D. Bae, "A two-step machine learning approach for crop disease detection: an application of GAN and UAV technology." arXiv, Sep. 18, 2021. Accessed: Jun. 19, 2022. [Online]. Available: <http://arxiv.org/abs/2109.11066>
70. W. Albattah, A. Javed, M. Nawaz, M. Masood, and S. Albahli, "Artificial Intelligence-Based Drone System for Multiclass Plant Disease Detection Using an Improved Efficient Convolutional Neural Network," *Frontiers in Plant Science*, vol. 13, Jun. 2022, doi: 10.3389/fpls.2022.808380.

71. S. Mishra, R. Sachan, and D. Rajpal, "Deep Convolutional Neural Network based Detection System for Real-time Corn Plant Disease Recognition," *Procedia Computer Science*, vol. 167, pp. 2003–2010, Jan. 2020, doi: 10.1016/j.procs.2020.03.236.
72. M. D. Bah, A. Hafiane, and R. Canals, "Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images," *Remote Sensing*, vol. 10, no. 11, Art. no. 11, Nov. 2018, doi: 10.3390/rs10111690.
73. Y. Zheng, E. Sarigul, G. Panicker, and D. Stott, "Vineyard LAI and canopy coverage estimation with convolutional neural network models and drone pictures," in *Sensing for Agriculture and Food Quality and Safety XIV*, SPIE, Jun. 2022, pp. 29–38. doi: 10.1117/12.2620100.
74. R. Yang *et al.*, "A Multi-Source Data Fusion Decision-Making Method for Disease and Pest Detection of Grape Foliage Based on ShuffleNet V2," *Remote Sensing*, vol. 13, no. 24, Art. no. 24, Jan. 2021, doi: 10.3390/rs13245102.
75. S. Briechle, P. Krzystek, and G. Vosselman, "CLASSIFICATION OF TREE SPECIES AND STANDING DEAD TREES BY FUSING UAV-BASED LIDAR DATA AND MULTISPECTRAL IMAGERY IN THE 3D DEEP NEURAL NETWORK POINTNET++," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. V-2–2020, pp. 203–210, Aug. 2020, doi: 10.5194/isprs-annals-V-2-2020-203-2020.
76. D. Aiger, B. Allen, and A. Golovinskiy, "Large-Scale 3D Scene Classification With Multi-View Volumetric CNN," arXiv, Dec. 26, 2017. Accessed: Jun. 07, 2022. [Online]. Available: <http://arxiv.org/abs/1712.09216>
77. M. D. Bah, A. Hafiane, and R. Canals, "CRoWNet: Deep Network for Crop Row Detection in UAV Images," *IEEE Access*, vol. 8, pp. 5189–5200, 2020, doi: 10.1109/ACCESS.2019.2960873.
78. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv, May 18, 2015. Accessed: Jun. 19, 2022. [Online]. Available: <http://arxiv.org/abs/1505.04597>
79. B. T. Kitano, C. C. T. Mendes, A. R. Geus, H. C. Oliveira, and J. R. Souza, "Corn Plant Counting Using Deep Learning and UAV Images," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2019, doi: 10.1109/LGRS.2019.2930549.
80. Z. Lin and W. Guo, "Sorghum Panicle Detection and Counting Using Unmanned Aerial System Images and Deep Learning," *Frontiers in Plant Science*, vol. 11, 2020, Accessed: Jun. 01, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpls.2020.534853>
81. L. El Hoummaidi, A. Larabi, and K. Alam, "Using unmanned aerial systems and deep learning for agriculture mapping in Dubai | Elsevier Enhanced Reader." <https://reader.elsevier.com/reader/sd/pii/S240584402102257X?token=08CBD9EDFD77D8950813E1351450D76E494A52920C53A724FFC4763B2D03C6A459F63D6C1C254EF04EBB26AA5515DACC&originRegion=eu-west-1&originCreation=20220601233300> (accessed Jun. 02, 2022).
82. R. Doha, M. Al Hasan, S. Anwar, and V. Rajendran, "Deep Learning based Crop Row Detection with Online Domain Adaptation," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Virtual Event Singapore: ACM, Aug. 2021, pp. 2773–2781. doi: 10.1145/3447548.3467155.
83. T. Zhang, Z. Yang, Z. Xu, and J. Li, "Wheat Yellow Rust Severity Detection by Efficient DF-UNet and UAV Multispectral Imagery," *IEEE Sensors Journal*, vol. 22, no. 9, pp. 9057–9068, May 2022, doi: 10.1109/JSEN.2022.3156097.
84. S. Tsuichihara *et al.*, "Drone and GPS Sensors-Based Grassland Management Using Deep-Learning Image Segmentation," in *2019 Third IEEE International Conference on Robotic Computing (IRC)*, Feb. 2019, pp. 608–611. doi: 10.1109/IRC.2019.00123.
85. M.-D. Yang, J. G. Boubin, H. P. Tsai, H.-H. Tseng, Y.-C. Hsu, and C. C. Stewart, "Adaptive autonomous UAV scouting for rice lodging assessment using edge computing with deep learning EDANet," *Computers and Electronics in Agriculture*, vol. 179, p. 105817, Dec. 2020, doi: 10.1016/j.compag.2020.105817.
86. J. Weyler, F. Magistri, P. Seitz, J. Behley, and C. Stachniss, "In-Field Phenotyping Based on Crop Leaf and Plant Instance Segmentation," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2022, pp. 2968–2977. doi: 10.1109/WACV51458.2022.00302.
87. Y. Guo *et al.*, "Plant Disease Identification Based on Deep Learning Algorithm in Smart Farming," *Discrete Dynamics in Nature and Society*, vol. 2020, pp. 1–11, Aug. 2020, doi: 10.1155/2020/2479172.
88. J. Torres-Sánchez, F. J. Mesas-Carrascosa, F. M. Jiménez-Brenes, A. I. de Castro, and F. López-Granados, "Early Detection of Broad-Leaved and Grass Weeds in Wide Row Crops Using Artificial Neural Networks and UAV Imagery," *Agronomy*, vol. 11, no. 4, Art. no. 4, Apr. 2021, doi: 10.3390/agronomy11040749.
89. X. Zhang, N. Li, L. Ge, X. Xia, and N. Ding, "A Unified Model for Real-Time Crop Recognition and Stem Localization Exploiting Cross-Task Feature Fusion," in *2020 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, Sep. 2020, pp. 327–332. doi: 10.1109/RCAR49640.2020.9303270.
90. N. Li, X. Zhang, C. Zhang, H. Guo, Z. Sun, and X. Wu, "Real-Time Crop Recognition in Transplanted Fields With Prominent Weed Growth: A Visual-Attention-Based Approach," *IEEE Access*, vol. 7, pp. 185310–185321, 2019, doi: 10.1109/ACCESS.2019.2942158.

91. C.-J. Chen, Y.-Y. Huang, Y.-S. Li, Y.-C. Chen, C.-Y. Chang, and Y.-M. Huang, "Identification of Fruit Tree Pests With Deep Learning on Embedded Drone to Achieve Accurate Pesticide Spraying," *IEEE Access*, vol. 9, pp. 21986–21997, 2021, doi: 10.1109/ACCESS.2021.3056082.
92. Z. Qin, W. Wang, K.-H. Dammer, L. Guo, and Z. Cao, "A Real-time Low-cost Artificial Intelligence System for Autonomous Spraying in Palm Plantations." arXiv, Mar. 06, 2021. Accessed: Jun. 15, 2022. [Online]. Available: <http://arxiv.org/abs/2103.04132>
93. A. I. B. Parico and T. Ahamed, "An Aerial Weed Detection System for Green Onion Crops Using the You Only Look Once (YOLOv3) Deep Learning Algorithm," *Engineering in Agriculture, Environment and Food*, vol. 13, no. 2, pp. 42–48, 2020, doi: 10.37221/eaef.13.2\_42.
94. C. Rui, G. Youwei, Z. Huafei, and J. Hongyu, "A Comprehensive Approach for UAV Small Object Detection with Simulation-based Transfer Learning and Adaptive Fusion." arXiv, Sep. 04, 2021. Accessed: Jun. 15, 2022. [Online]. Available: <http://arxiv.org/abs/2109.01800>
95. A. I. B. Parico and T. Ahamed, "Real Time Pear Fruit Detection and Counting Using YOLOv4 Models and Deep SORT," *Sensors*, vol. 21, no. 14, p. 4803, Jul. 2021, doi: 10.3390/s21144803.
96. T. Jintasuttisak, E. Edirisinghe, and A. Elbattay, "Deep neural network based date palm tree detection in drone imagery," *Computers and Electronics in Agriculture*, vol. 192, p. 106560, Jan. 2022, doi: 10.1016/j.compag.2021.106560.
97. Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOv3-Dense." [https://www.researchgate.net/publication/355887041\\_Detection\\_of\\_Apple\\_Lesions\\_in\\_Orchards\\_Based\\_on\\_Deep\\_Learning\\_Methods\\_of\\_CycleGAN\\_and\\_YOLOv3-Dense](https://www.researchgate.net/publication/355887041_Detection_of_Apple_Lesions_in_Orchards_Based_on_Deep_Learning_Methods_of_CycleGAN_and_YOLOv3-Dense) (accessed Jun. 20, 2022).
98. W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0\_2.
99. A. N. Veeranampalayam Sivakumar *et al.*, "Comparison of Object Detection and Patch-Based Classification Deep Learning Models on Mid- to Late-Season Weed Detection in UAV Imagery," *Remote Sensing*, vol. 12, no. 13, p. 2136, Jul. 2020, doi: 10.3390/rs12132136.
100. M. F. Ridho and Irwan, "Strawberry Fruit Quality Assessment for Harvesting Robot using SSD Convolutional Neural Network," in *2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Oct. 2021, pp. 157–162. doi: 10.23919/EECSI53397.2021.9624311.
101. A. Ammar, A. Koubaa, and B. Benjdira, "Deep-Learning-Based Automated Palm Tree Counting and Geolocation in Large Farms from Aerial Geotagged Images," *Agronomy*, vol. 11, no. 8, Art. no. 8, Aug. 2021, doi: 10.3390/agronomy11081458.
102. W.-H. Su *et al.*, "Automatic Evaluation of Wheat Resistance to Fusarium Head Blight Using Dual Mask-RCNN Deep Learning Frameworks in Computer Vision." [https://www.researchgate.net/publication/347793558\\_Automatic\\_Evaluation\\_of\\_Wheat\\_Resistance\\_to\\_Fusarium\\_Head\\_Blight\\_Using\\_Dual\\_Mask-RCNN\\_Deep\\_Learning\\_Frameworks\\_in\\_Computer\\_Vision](https://www.researchgate.net/publication/347793558_Automatic_Evaluation_of_Wheat_Resistance_to_Fusarium_Head_Blight_Using_Dual_Mask-RCNN_Deep_Learning_Frameworks_in_Computer_Vision) (accessed Jun. 20, 2022).
103. M. D. Yang, H. H. Tseng, Y. C. Hsu, and W. C. Tseng, "Real-time Crop Classification Using Edge Computing and Deep Learning," in *2020 IEEE 17th Annual Consumer Communications Networking Conference (CCNC)*, Jan. 2020, pp. 1–4. doi: 10.1109/CCNC46108.2020.9045498.
104. A. Menshchikov *et al.*, "Real-Time Detection of Hogweed: UAV Platform Empowered by Deep Learning," *IEEE Transactions on Computers*, vol. 70, no. 8, pp. 1175–1188, Aug. 2021, doi: 10.1109/TC.2021.3059819.
105. J. Weyler, J. Quakernack, P. Lottes, J. Behley, and C. Stachniss, "Joint Plant and Leaf Instance Segmentation on Field-Scale UAV Imagery," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3787–3794, Apr. 2022, doi: 10.1109/LRA.2022.3147462.
106. P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss, "Robust joint stem detection and crop-weed classification using image sequences for plant-specific treatment in precision farming." [https://www.researchgate.net/publication/335117645\\_Robust\\_joint\\_stem\\_detection\\_and\\_crop-weed\\_classification\\_using\\_image\\_sequences\\_for\\_plant-specific\\_treatment\\_in\\_precision\\_farming](https://www.researchgate.net/publication/335117645_Robust_joint_stem_detection_and_crop-weed_classification_using_image_sequences_for_plant-specific_treatment_in_precision_farming) (accessed Jun. 07, 2022).
107. D. Su, Y. Qiao, H. Kong, and S. Sukkarieh, "Real time detection of inter-row ryegrass in wheat farms using deep learning," *Biosystems Engineering*, vol. 204, pp. 198–211, Apr. 2021, doi: 10.1016/j.biosystemseng.2021.01.019.
108. A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Jun. 18, 2022. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
109. A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, Jun. 03, 2021. doi: 10.48550/arXiv.2010.11929.
110. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers." arXiv, May 28, 2020. doi: 10.48550/arXiv.2005.12872.

111. H.-T. Thai, N.-Y. Tran-Van, and K.-H. Le, "Artificial Cognition for Early Leaf Disease Detection using Vision Transformers," in *2021 International Conference on Advanced Technologies for Communications (ATC)*, Oct. 2021, pp. 33–38. doi: 10.1109/ATC52653.2021.9598303.
112. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," p. 8.
113. "Cassava Leaf Disease Classification." <https://kaggle.com/competitions/cassava-leaf-disease-classification> (accessed Jun. 18, 2022).
114. K. Karila *et al.*, "Estimating Grass Sward Quality and Quantity Parameters Using Drone Remote Sensing with Deep Neural Networks," *Remote Sensing*, vol. 14, no. 11, Art. no. 11, Jan. 2022, doi: 10.3390/rs14112692.
115. S. Dersch, A. Schottl, P. Krzystek, and M. Heurich, "NOVEL SINGLE TREE DETECTION BY TRANSFORMERS USING UAV-BASED MULTISPECTRAL IMAGERY - ProQuest." <https://www.proquest.com/openview/228f8f292353d30b26ebcdd38372d40d/1?pq-origsite=gscholar&cbl=2037674> (accessed Jun. 15, 2022).
116. G. Chen and Y. Shang, "Transformer for Tree Counting in Aerial Images," *Remote Sensing*, vol. 14, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/rs14030476.
117. W. Liu, M. Salzmann, and P. Fua, "Context-Aware Crowd Counting." arXiv, Apr. 15, 2019. Accessed: Jun. 19, 2022. [Online]. Available: <http://arxiv.org/abs/1811.10452>
118. X. Zhang *et al.*, "The self-supervised spectral-spatial attention-based transformer network for automated, accurate prediction of crop nitrogen status from UAV imagery," arXiv, arXiv:2111.06839, Feb. 2022. doi: 10.48550/arXiv.2111.06839.
119. L. F. S. Coletta, D. C. de Almeida, J. R. Souza, and R. L. Manzione, "Novelty detection in UAV images to identify emerging threats in eucalyptus crops," *Computers and Electronics in Agriculture*, vol. 196, p. 106901, May 2022, doi: 10.1016/j.compag.2022.106901.
120. C. Silva *et al.*, "Imputation of Individual Longleaf Pine (*Pinus palustris* Mill.) Tree Attributes from Field and LiDAR Data: Canadian Journal of Remote Sensing: Vol 42, No 5." <https://www.tandfonline.com/doi/full/10.1080/07038992.2016.1196582> (accessed Jun. 20, 2022).
121. P. Bosilj, E. Aptoula, T. Duckett, and G. Cielniak, "Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture," *Journal of Field Robotics*, Mar. 2019, doi: 10.1002/rob.21869.
122. S. Khan, M. Tufail, M. T. Khan, Z. A. Khan, J. Iqbal, and M. Alam, "A novel semi-supervised framework for UAV based crop/weed classification," *PLOS ONE*, vol. 16, no. 5, p. e0251008, May 2021, doi: 10.1371/journal.pone.0251008.
123. B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks," *Remote Sensing*, vol. 11, no. 11, Art. no. 11, Jan. 2019, doi: 10.3390/rs11111309.
124. Z. Alom, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," p. 12.
125. M. Fawakherji, C. Potena, I. Prevedello, A. Pretto, D. D. Bloisi, and D. Nardi, "Data Augmentation Using GANs for Crop/Weed Segmentation in Precision Farming," in *2020 IEEE Conference on Control Technology and Applications (CCTA)*, Aug. 2020, pp. 279–284. doi: 10.1109/CCTA41146.2020.9206297.
126. M. Fawakherji, C. Potena, A. Pretto, D. D. Bloisi, and D. Nardi, "Multi-Spectral Image Synthesis for Crop/Weed Segmentation in Precision Farming," *Robotics and Autonomous Systems*, vol. 146, p. 103861, Dec. 2021, doi: 10.1016/j.robot.2021.103861.
127. V. Mazzia, A. Khaliq, F. Salvetti, and M. Chiaberge, "Real-Time Apple Detection System Using Embedded Systems With Hardware Accelerators: An Edge AI Application," *IEEE Access*, vol. 8, pp. 9102–9114, 2020, doi: 10.1109/ACCESS.2020.2964608.
128. Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, "Efficient Transformers: A Survey," *ACM Comput. Surv.*, vol. 55, no. 6, p. 109:1–109:28, Dec. 2022, doi: 10.1145/3530811.
129. M. Poli *et al.*, "Hyena Hierarchy: Towards Larger Convolutional Language Models." arXiv, Mar. 05, 2023. doi: 10.48550/arXiv.2302.10866.
130. J. J. Senecal, J. W. Sheppard, and J. A. Shaw, "Efficient Convolutional Neural Networks for Multi-Spectral Image Classification," in *2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary: IEEE, Jul. 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8851840.
131. Z. Yang, J. Li, X. Shi, and Z. Xu, "Dual flow transformer network for multispectral image segmentation of wheat yellow rust," in *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2022)*, SPIE, Dec. 2022, pp. 119–125. doi: 10.1117/12.2640930.
132. C. Tao *et al.*, "MSNet: multispectral semantic segmentation network for remote sensing images," *GIScience & Remote Sensing*, vol. 59, no. 1, pp. 1177–1198, Dec. 2022, doi: 10.1080/15481603.2022.2101728.
133. M. López and J. Alberto, "The use of multispectral images and deep learning models for agriculture: the application on Agave," Dec. 2022, Accessed: Mar. 12, 2023. [Online]. Available: <https://repositorio.tec.mx/handle/11285/650159>

134. B. Victor, Z. He, and A. Nibali, "A systematic review of the use of Deep Learning in Satellite Imagery for Agriculture." arXiv, Oct. 03, 2022. doi: 10.48550/arXiv.2210.01272.
135. P. Sarigiannidis, "Peach Tree Disease Detection Dataset." IEEE, Nov. 23, 2022. Accessed: Mar. 12, 2023. [Online]. Available: <https://ieee-dataport.org/documents/peach-tree-disease-detection-dataset>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.