Article

Not peer-reviewed version

Deep Learning-Based Weed Detection and Classification in Wheat Fields from UAV Imagery

Sarangerel Jarantaibaatar , Md. Shiful Islam , Yago Diez * , Maximo Larry Lopez Caceres * , Myagmarjav Indra , Tobias Leidemer , Vladislav Bukin , Shinsuke Konno , Shinebayar Turbat , Batbileg Bayaraa , Jun Yokoyama , Nakamura Atsushi , Burmaa Chuluunbat , Leonardo Silvestre , Federico Giovanni Nicola

Posted Date: 7 April 2026

doi: 10.20944/preprints202604.0402.v1

Keywords: deep learning; instance segmentation; mask R-CNN; precision agriculture; weed detection; YOLOv8; wheat production

Article

Deep Learning-Based Weed Detection and Classification in Wheat Fields from UAV Imagery

Sarangerel Jarantaibaatar [1], Md. Shiful Islam [1], Yago Diez [2,*], Maximo Larry Lopez Caceres [1,*], Myagmarjav Indra [3], Tobias Leidemer [1], Vladislav Bukin [1], Shinsuke Konno [1], Shinebayar Turbat [3], Batbileg Bayaraa [3], Jun Yokoyama [2], Nakamura Atsushi [4], Burmaa Chuluunbat [1], Leonardo Silvestre [1] and Federico Giovanni Nicola [1]

[1] Faculty of Agriculture, Yamagata University, 1-23 Wakaba-machi, Tsuruoka, Yamagata 997- 8555, Japan
[2] Faculty of Science, Yamagata University, 1-4-12 Kojirakawa-machi, Yamagata 990-8560, Japan
[3] Faculty of Agro-ecology, Mongolian University of Life Sciences, District 22 Zaisan, Ulaanbaatar 17024, Mongolia
[4] Faculty of Humanities and Social Sciences, Yamagata University, 1-4-12 Kojirakawa-machi, Yamagata 990-8560, Japan
* Correspondence: yago@sci.kj.yamagata-u.ac.jp (Y.D.); larry@tds1.tr.yamagata-u.ac.jp (M.L.L.C.)

Abstract

Weed infestation significantly threatens crop productivity and quality, highlighting the need for accurate and scalable monitoring approaches. Recent advances in unmanned aerial vehicle (UAV) remote sensing and deep learning provide promising tools for field-scale weed detection. This study evaluates and compares two state-of-the-art instance segmentation models, Mask R-CNN and YOLOv8, for species-level weed detection in wheat fields under Mongolian agro-ecological conditions. The experiment was conducted in a 4 ha wheat field in Tuv Province, Mongolia, using high-resolution RGB imagery acquired from UAV flights in July 2025. Three dominant weed species were annotated and analyzed. Model performance was evaluated using mAP@0.5:0.95, Precision, Recall, F1-score, and mask IoU. At IoU thresholds of 0.25 and 0.5, both models demonstrated moderate detection performance (IoU = 0.25: Precision 0.49–0.76, Recall 0.20–0.77, F1-score 0.32–0.75; IoU = 0.5: Precision 0.42–0.67, Recall 0.18–0.75, F1-score 0.28–0.69), with variation among weed species. Mask R-CNN achieved higher Recall and more precise boundary delineation, improving weed coverage estimation, whereas YOLOv8 provided faster inference (≈11 ms per image, ~90 FPS) and higher precision, making it more suitable for large-area and near-real-time monitoring. These findings demonstrate the potential of UAV-based instance segmentation for weed detection in Mongolia and provide practical guidance for model selection in precision agriculture applications.

Keywords: deep learning; instance segmentation; mask R-CNN; precision agriculture; weed detection; YOLOv8; wheat production

---

1. Introduction

Climate change and rapid population growth pose significant challenges to global food security, particularly in the agricultural sector. According to the Food and Agriculture Organization (FAO) global food production must increase by approximately 70% by 2050 to meet the demands of a growing population, projected to reach nearly nine billion [1]. Achieving this target requires minimizing yield losses in staple crops, including cereals, wheat, and vegetables. Globally, wheat (*Triticum aestivum* L.) is cultivated in over 220 million hectares with a production of nearly 799 million tons in 2023 [2].

In Mongolia, wheat is the primary food crop, accounting for 1.6% of national GDP and 8.9% of agricultural value added [3]. As of 2025, wheat planting covered 232.8 thousand hectares, representing 75.1% of the planted cropping area [4]. However, wheat production is increasingly threatened by yield losses caused by environmental stressors, limited labor and machinery availability, extreme climatic conditions, and mainly by widespread weed infestation. Weeds are a major biotic factor reducing wheat yield, causing economic losses of 40–50% at the field level and up to 43% globally if unmanaged, as they compete with crops for nutrients, water, and sunlight and also host insect pests and diseases that degrade crop quality and productivity [5]. Conventional blanket herbicide application, although effective, often leads to excessive chemical use, increased production costs, environmental pollution, and food safety concerns [6].

Traditional weed monitoring in Mongolia relies on labor-intensive field surveys based on sampling protocols established in the 1980s [7]. These methods involve sparse ground sampling (e.g., 1 m^2 plots per 10–25 ha), which often fails to capture the spatial heterogeneity of weed distribution and leads to inaccurate weed density estimation. Given Mongolia's vast agricultural landscapes, limited workforce, and harsh climatic conditions, predominantly arid and fragile soil conditions, such approaches are increasingly inadequate for timely and accurate weed management. Therefore, transitioning toward automated, data-driven, and scalable weed monitoring technologies is imperative. Recent advances in unmanned aerial vehicles (UAVs), and Deep learning (DL) are significantly transforming weed detection and mapping methodologies. As drones can generate a large number of high-resolution aerial images of plants, these images require automated and accurate analysis so that farmers can obtain the necessary knowledge about their crops [8]. Utilizing drone imagery for weed detection offers a strong advantage over other field-based or imaging methods because drones can hover at low altitudes and capture high-resolution images with enhanced spatial detail [9].

These inherent shortcomings of existing weed management and monitoring approaches have directly boosted the rapid development and adoption of UAV technology in precision agriculture [10]. UAVs are commonly used to capture aerial imagery, and can easily satisfy the desire to obtain high-resolution data, making UAV-based imagery important for DL-based weed detection and classification in different agricultural fields [11]. From this, it can be seen that using UAV and DL together is an effective way to monitor weed growth, development, and distribution. DL particularly focuses on the depth of the model structure, emphasizes the importance of feature learning, and proposes various techniques to learn more and higher-level features more effectively and quickly [12]. Convolutional Neural Networks (CNNs) are a distinct type of DL model recognized for their high effectiveness in image recognition and classification tasks [13]. DL has made it possible to monitor crops, predict yield, detect diseases, manage soil health, forecast weather, optimize supply chains, support genomics and breeding, practice precision agriculture, and analyze markets, greatly transforming the agricultural sector [14]. Recent advancements in UAV technology combined with DL have revolutionized weed monitoring, offering superior accuracy, scalability, and operational feasibility compared to traditional methods [15].

More specifically, DL network architectures, including Mask R-CNN [16], and YOLO [17], are increasingly utilized for a range of applications in agriculture. DL–based instance segmentation models, such as Mask R-CNN and YOLOv8, enable end-to-end detection and pixel-level delineation of weeds, overcoming the limitations of traditional machine vision approaches that require manual feature extraction [18]. However, larger and more complex image datasets present significant challenges, as computational costs increase dramatically and annotation requires labor-intensive work by trained experts, which has driven the development of new generations of DL architectures with improved network topologies that enable faster parallel processing and more effective information extraction from input images [19]. Therefore, this study aims to evaluate and compare the performance of Mask R-CNN and YOLOv8 for UAV-based weed detection and mapping in wheat fields.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0402.v1

2. Materials and Methods

2.1. Study Site

The study was conducted in Bornuur district, Tuv Province, northern Mongolia, which is one of the country's major wheat-producing regions. The study area is located approximately 115 km north of the capital Ulaanbaatar and lies at an elevation ranging from 1,000 to 1,500 m.a.s.l. Geographically, the area belongs to the transitional forest–steppe zone and is characterized by favorable conditions for both crop production and livestock husbandry. The study area, Bornuur, is characterized by a continental semi-arid climate with cold, dry winters and warm summers. Air temperature typically ranges from −28 °C to 25 °C. The mean annual air temperature ranges between −2 and 0 °C. Annual precipitation varies between 250 and 300 mm, with more than 70% occurring during the summer months. This pronounced seasonal concentration of precipitation strongly influences crop growth patterns and weed emergence dynamics in the region. The site area is part of the central agricultural zone of Mongolia and belongs to the long-term soil management and crop rotation experimental field of the National University of Mongolia's agricultural research and training center "Nart Center" (Figure 1).

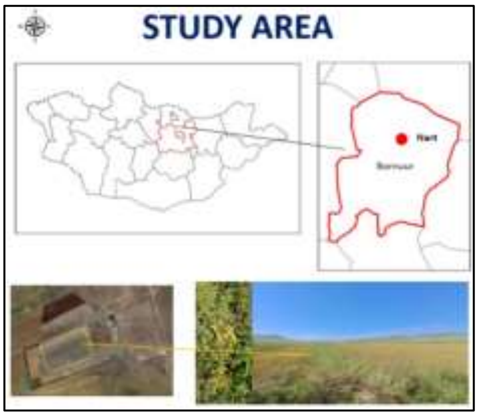
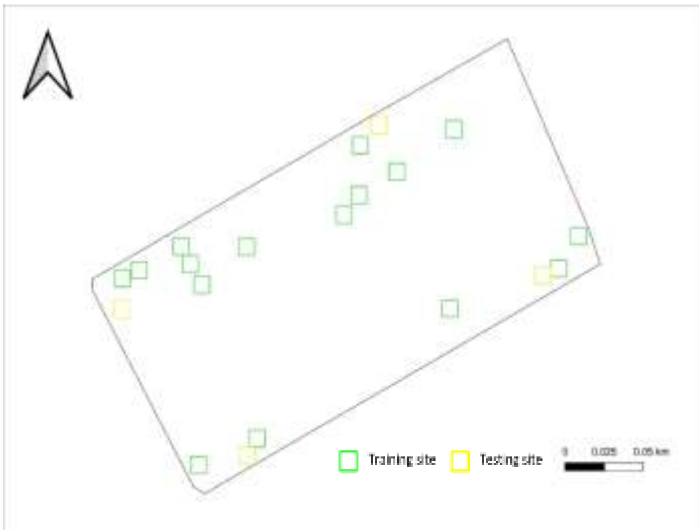


Figure 1. Location of the study area in Bornuur soum, Tuv Province, northern Mongolia, /48°40′45.7″ N, 106°16′01.8″ E/.

To ensure representative sampling and minimize spatial bias, the field was surveyed following a zigzag transect, and subsequently subdivided into smaller plots. This approach facilitated systematic coverage of heterogeneous weed distributions and enabled the preparation of a robust training dataset for deep learning–based weed detection and segmentation (Figure 2).



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0402.v1

Figure 2. Study sites, (Each with a size of 10m x 10m, (100m^2)).

2.2. Weed Research

A substantial proportion of the 689 species identified in Mongolia [20], occur as weeds in agricultural lands, limits crop production. In the central agricultural region of Mongolia, most of the 361 vascular plant species have been recognized as arable land weeds [21]. At the provincial scale, surveys conducted in Tuv Province reported the presence of 21 weed species belonging to 19 genera and 9 families across 13,668 ha of cultivated land. Furthermore, a broader assessment in the central agricultural region-including Tuv, Selenge, Darkhan-Uul, and Bulgan provinces identified 66 weed species in wheat, potato, vegetable, and fruit-growing areas [22]. At the Nart experimental site, weed flora surveys identified 18 weed species where annual weeds accounted for 61.1% of the total species, followed by perennials (33.3%) and biennials (5.6%) [7].

2.3. Deep Learning Model

2.3.1. Model Mask R-CNN

Mask R-CNN is a two-stage instance segmentation framework that extends Faster R-CNN by incorporating a parallel branch for predicting segmentation masks for each RoI at the pixel level [16]. To extract spatially consistent features and guarantee accurate segmentation borders, this branch combines ROI Align [23]. Due to its high segmentation accuracy, Mask R-CNN has been widely applied in precision agriculture and remote sensing applications [18]. In this study, it was adopted as an accuracy-oriented baseline model for weed detection and segmentation.

2.3.2. Model Ultralytics YOLOv8

Recently, YOLOv8 has emerged as a leading one-stage detector due to its balance between accuracy and real-time speed; however, challenging cases where objects are obscured, small, or visually ambiguous remain a bottleneck, requiring more refined representation learning and robust feature extraction for reliable detection [24]. Then YOLO (You Only Look Once) came along, changing things up by approaching object detection as a single-step regression task [25]. The YOLO models excel in real-time applications by delivering a high frame rate without significant compromise on detection accuracy [26]. Its anchor-free design and unified detection–segmentation pipeline make it particularly suitable for UAV-based agricultural monitoring [18]. Contemporary literature confirms that advanced deep learning frameworks, specifically the YOLOv8 series, achieve precise pixel-level delineation of weeds and crops when processing high-resolution drone orthomosaics [27].

2.3.3. Comparison of Model Characteristics

Recent UAV-based agricultural studies provide empirical evidence supporting the selection of YOLOv8 and Mask R-CNN for weed instance segmentation. For example, In the instance segmentation task of apple orchards, YOLOv8 demonstrated superior performance to Mask R-CNN, specifically, YOLOv8 achieved 92.9% Precision and 97% Recall in a single-class segmentation task, while Mask R-CNN achieved 84.7% Precision and 88% Recall [28]. However, there is currently no research that has compared the capabilities of Mask R-CNN and YOLOv8 in segmenting weeds in UAV images [29]. The comparison between the models demonstrated that YOLOv8 performed significantly better than Mask R-CNN (Detectron2) and U-Net in three of the four performance metrics. Therefore, the YOLOv8 model and its variants presented the highest mAP and Recall results, as well as the highest accuracy result [30]. In this study, we proposed WeedSense, a novel multi-task learning architecture for comprehensive weed analysis that simultaneously performs semantic segmentation, height estimation, and growth stage classification using RGB imagery [31]. Surpassing earlier versions and detectors such as Faster R-CNN (e.g., 1.3 ms vs. 54 ms GPU latency) by utilizing a new backbone, an anchor-free detection head, and efficient C2f modules [32] (Table 1).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0402.v1

Table 1. Structural dimension of Mask R-CNN and YOLOv8.

	Mask R-CNN	YOLOv8
Architecture	Two-stage, region-based	Single—stage, anchor—free
Strength	Pixel-level accuracy	Speed + accuracy balance
Use case	High-precision segmentation	Real—time UAV applications

2.4. Models application

The comparative evaluation of Mask R-CNN and YOLOv8 models enable a systematic analysis of the trade-offs between real-time performance and segmentation accuracy under identical datasets and experimental conditions. The Model Training Configuration section presents the preparation of the training dataset and the parameters used for model training (Table 2).

Table 2. Model Training Configuration.

Parameter	Value/Unit	Value/Unit
Model	Mask R-CNN	YOLOv8
Input size	512x512	512x512
Epochs	20	200
Validation parameter	• Prediction loss • Objectness loss • Class loss	• Box loss • Objectness loss • Class loss

An experimental framework was developed to compare the Mask R-CNN and YOLOv8 models in terms of the metrics used for weed detection and the approaches applied for detailed performance evaluation (Figure 3).

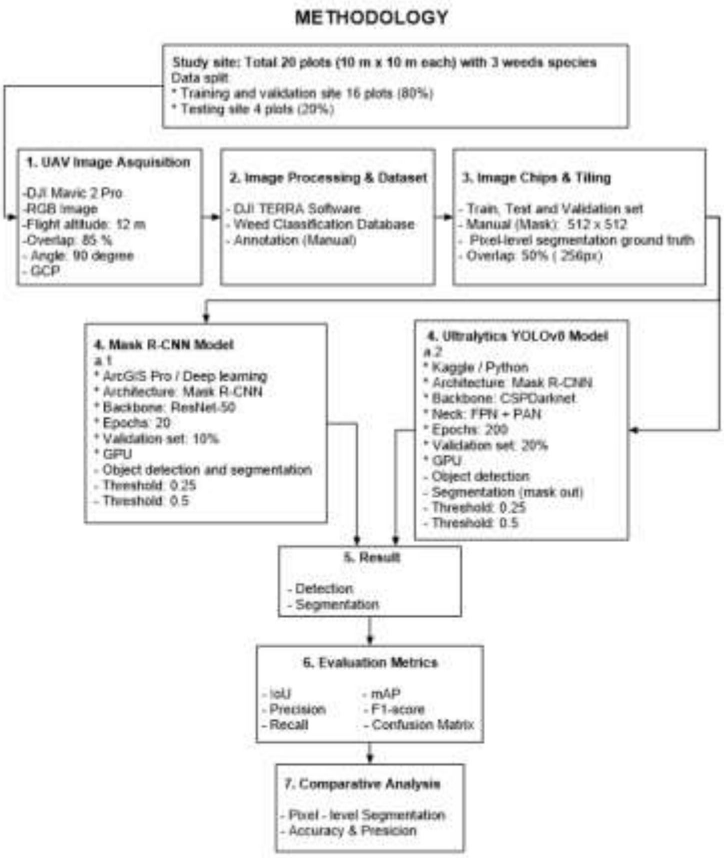


Figure 3. Workflow of the Mask R-CNN and YOLOv8 Models.

2.5. Evaluation Metrics

In this study, model performance was evaluated using standard classification and object detection metrics, including True Positive (TP), False Positive (FP), and False Negative (FN). Based on TP, FP, and FN were calculated to assess detection accuracy. Precision (positive predictive value) measures the proportion of correctly detected objects among all predicted positives, reflecting the model's ability to avoid false detections. Recall (sensitivity), on the other hand, represents the proportion of correctly detected objects among all ground-truth positives, indicating the model's capability to identify all relevant instances [33]. To summarize model performance across varying confidence thresholds, Average Precision (AP) (Eq. 2) was employed. The mean Average Precision (mAP) is obtained by averaging AP values across all object classes, providing a comprehensive measure of the overall detection performance [34]. mAP50 is the mAP calculated at an IoU (Eq. 5) threshold of 0.50, and mAP50-95 is the mAP averaged over IoU thresholds ranging from 0.50 to 0.95, with an increment of 0.05 [35]. The F1-score (Eq. 4) is the harmonic mean of Precision and Recall, ranging from 0 to 1, where higher values indicate better performance. [36]. Recall quantifies how many actual positive instances are correctly identified. It is calculated as the proportion of accurately identified positive samples to all positive samples [37]. IoU is a metric that quantifies the overlap between the predicted bounding box and the ground truth bounding box relative to the area of their union [38].

$$P = \frac{TP}{TP+FP} \tag{1}$$

$$AP = \int_0^1 P(R)dR \tag{2}$$

$$R = \frac{TP}{TP+FN} \tag{3}$$

$$F1 = 2x\frac{PR}{P+R} \tag{4}$$

$$IoU_k = \frac{1}{K}\sum_{k=1}^{K} IoU_k \tag{5}$$

2.6. UAV Image Processing and Data Collection

The dataset used in this study was derived from UAV imagery acquired on 10 July 2025, combined with field survey data on weed occurrence and spatial distribution collected on the same date. Field observations conducted on the same day identified the presence of 11 weed species belonging to 9 families within the study area. Among these, the three most dominant weed species were selected for analysis. During the 2025 growing season, exceptionally dry conditions resulted in significantly reduced growth of both wheat and weeds. Due to the poor wheat stand, weed identification was comparatively easier during the survey period. A Mavic 2 Pro drone was used in this study (Table 3). Current UAV systems often rely on GPS coordinates for spraying without distinguishing between crops and weeds [39]. In addition, a high-precision GPS device was employed to ensure accurate coordinate measurements. The parameters are presented in the table below.

Table 3. Mavic 2 Pro Parameters.

Parameter	Unit
Model	Mavic 2Pro
Camera	Hasselblad L1D-20c
Sensor	1" CMOS, RGB
Resolution	5472×3648
Flight high	15m
Overlap	85%
Camera angle	90°
Pixel size	2.4 µm

 doi:10.20944/preprints202604.0402.v1

GSD	2.85 cm / pixel
GNSS	GPS+GLONASS

Drones equipped with object detection technologies can identify and follow both stationary and moving objects in real-time [40]. During the field survey phase on this time, UAV flights were conducted over an area of 4 ha. The UAV flight was conducted at an altitude of 15 m with a camera nadir angle of 90°, 85% overlapping, to acquire high-resolution RGB imagery, resulting in a Ground Sample Distance of 0.285 cm/pixel. A total of eight GCPs were distributed across the field to ensure accurate georeferencing. In addition, in situ weed surveys were conducted, with weed species visually identified, recorded, and documented.

The balanced allocation of the study area is a critical factor influencing the overall research outcomes. Proper proportional distribution among these subsets ensures that the model is trained on sufficient data while allowing robust and unbiased performance evaluation. *L. buriatica*, *N. pulla*, *A. scoparia* species were selected not only for being the document species within the study area, but also because their bright yellow flowers and dark green leaves were easier to detect (Figure 4).

Figure 4. Data Collection species of weed, (**a**) *Linaria buriatica Turcz,* (**b**) *Neneo pulla (L.) DC.*, (**c**) *Artemisia scoparia Walds.*

During dataset annotation, the experimental area was divided into sites of equal size. Specifically, 10 m × 10 m were selected. 16 plots for training dataset and 4 plots for testing dataset (Table 4).

Table 4. Data set annotation per weed species divided into Training and Testing.

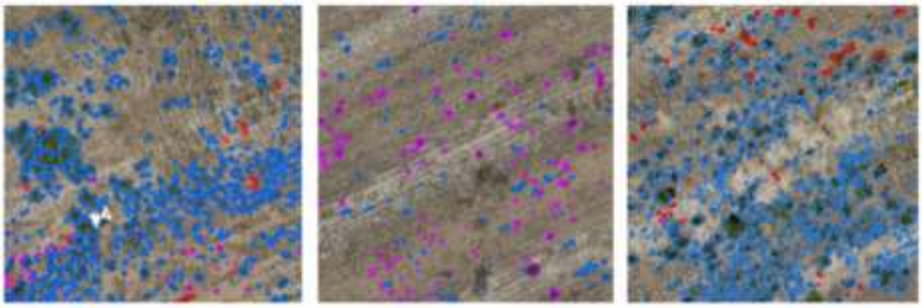
Species of weed	Train data	Test data
L. buriatica	3476	1211
N. pulla	749	259
A. scoparia	491	122
Total	4716	1592

 doi:10.20944/preprints202604.0402.v1

2.7. Data Processing-Data Base

During the image processing stage, raw UAV imagery was processed using DJI Terra software (SZ DJI Technology Co., Ltd., Shenzhen, China) to generate high-resolution RGB orthomosaic images.

The RGB data were imported into QGIS, where weeds and their spatial distribution were manually annotated and converted into vector-based datasets for DL training. The annotated weed were classified by species and organized into a structured database, with auxiliary attribute information recorded and numerical class labels assigned. Variability in weed appearance and partial occlusion introduced minor annotation inconsistencies; however, training the models on heterogeneous and visually complex samples is expected to enhance their generalization capability under real-world field conditions. Although the manual annotation process is inherently subjective, restricting labels to confidently identifiable weed instances improves dataset robustness and reduces the risk of systematic bias. This trade-off between annotation completeness and label reliability strengthens the practical applicability of the trained models, particularly for UAV-based weed mapping in complex agricultural environments. In some cases, discriminating weeds from RGB imagery was not easy, particularly when species exhibited similar morphological characteristics or grew intermingled or overlapped with other vegetation (Figure 5).



(a) (b) (c)

Figure 5. Annotation process of weeds, (**a**) Linaria buriatica Turcz, (**b**) Neneo pulla (L.) DC., (**c**) Artemisia scoparia Walds.

Weed detection using Mask R-CNN was performed with the DL tools implemented in ArcGIS Pro sowtware (Esri, Inc.LLC, Redlands, CA, USA). Training data were prepared by tiling the original orthomosaic images into patches of 512 × 512 pixels with a stride of 256 × 256 pixels, overlap 80%. Instance-level segmentation masks were manually generated, and the corresponding metadata were exported following the Mask R-CNN mask format. The Mask R-CNN model adopts a two-stage instance segmentation architecture, consisting of a region proposal network (RPN) followed by a classification and mask prediction stage. A ResNet-50 backbone was employed for hierarchical feature extraction. Model training was conducted on the training site using a validation split of 10%, and a fixed chip size of 512 × 512 pixels. After training, the optimized Mask R-CNN model was applied to the testing site to automatically detect and classify weed species. In parallel, weed detection was conducted using the YOLOv8 instance segmentation framework implemented in Python on the Kaggle platform. Polygon-based instance segmentation annotations were generated and converted into the YOLOv8 segmentation label format. Model training was performed using pre-trained YOLOv8 segmentation weights with 200 epochs, a batch size of 8 image patches, 4 worker threads for data loading, and a validation split of 20%. The trained model was subsequently applied to the testing plots to automatically detect, segment, and classify weed species.

 doi:10.20944/preprints202604.0402.v1

3. Results

3.1. IoU-Based Detection Performance of Mask R-CNN

The performance of the Mask R-CNN model for weed detection and instance segmentation was evaluated at IoU thresholds of 0.25 and 0.5 using Precision, Recall, F1-score, and average Precision (AP). At an IoU threshold of 0.25, *L. buriatica* achieved the highest detection performance among all species. The model obtained a Precision of 0.73, Recall of 0.77, and an F1-score of 0.75, with an AP of 0.69. A total of 935 true positives were correctly detected, demonstrating a detection accuracy of 77% for this species under the low overlap threshold. For *N. pulla*, Mask R-CNN demonstrated a Recall-oriented detection behavior. At IoU = 0.25, the model achieved a Recall of 0.75 and a Precision of 0.56, resulting in an F1-score of 0.64 and an AP of 0.55. Although a large proportion of instances were successfully detected, the relatively high number of false positives (149) were observed. The weakest performance at IoU = 0.25 was observed for *A. scoparia*, and Precision remained high at 0.76, Recall was limited to 0.20, yielding an F1-score of 0.32 and an AP of 0.16, representing a detection rate of 10%. Across all species, Mask R-CNN achieved an overall Precision of 0.68, Recall of 0.57, F1-score of 0.57, and a mean AP (mAP) of 0.47 at IoU = 0.25. When the IoU threshold was increased to 0.5, detection performance declined across all species. *L. buriatica* remained the best-performing class, with Precision, Recall, and F1-score of 0.67, 0.71, and 0.69, respectively, and AP 0.54. The number of true positives decreased to 859, accompanied by an increase in false negatives. For *N. pulla*, Recall decreased to 0.75 while Precision remained relatively stable at 0.56, resulting in an F1-score of 0.62 and an AP of 0.53. *A. scoparia* again exhibited the weakest performance, with Recall further declining to 0.18 and AP dropping to 0.12, indicating persistent difficulty in detecting this species under stricter detection constraints. Overall, at IoU = 0.5, Mask R-CNN achieved a mean Precision of 0.63, Recall of 0.54, F1-score of 0.53, and mAP of 0.54. This section presents the Mask R-CNN predictions in relation to the training and testing data (Table 5).

Table 5. Results of the models identifying weeds.

Species of weed	Test site Annotation	Mask RCNN Prediction	YOLOv8 Prediction
L. buriatica	1211	1279	1203
N. pulla	259	342	179
A. scoparia	122	33	91
Total	1592	1654	1473

The results of the Mask R-CNN model at IoU = 0.25 demonstrated higher accuracy than IoU = 0.5 (Table 6).

Table 6. Metrics of Model-Mask R-CNN.

	Species of weed	Precision	Recall	F1 Score	AP	True Positive	False Positive	False Negative
IoU: 0.25	*L. buriatica*	0.73	0.77	0.75	0.68	938	341	261
	N. pulla	0.56	0.75	0.64	0.55	193	149	65
	A. scoparia	0.76	0.20	0.32	0.16	25	8	97
	All	0.68	0.57	0.57	0.62	1156	498	423
IoU: 0.5	*L. buriatica*	0.67	0.71	0.69	0.58	859	420	352
	N. pulla	0.55	0.72	0.62	0.53	187	155	72
	A. scoparia	0.67	0.18	0.28	0.12	22	11	100
	All	0.63	0.54	0.53	0.54	1068	586	524

3.2. Confusion Matrix Results of Mask R-CNN

To further analyze classification behavior, confusion matrices were examined at IoU thresholds of 0.25 and 0.5 using both normalized proportions and raw detection counts (Figures 6 and 7). At IoU = 0.25, the confusion matrix indicates that *L. Buriatica* achieved the highest correct classification rate, with approximately 78% of true *L. buriatica* samples correctly predicted. A total of 938 instances were correctly classified, confirming the strong performance observed in the IoU-based metrics. Some misclassification with *N. Pulla* and *A. Scoparia* was observed. For *N. pulla*, classification accuracy was relatively high, with 193 true positives. However, a noticeable proportion of samples was misclassified as *L. buriatica* or *A. scoparia*, reflecting visual similarity and class ambiguity. *A. scoparia* exhibited substantial misclassification, with only 25 correct predictions at IoU = 0.25. Most *A. scoparia* samples were either misclassified into other species or not detected, highlighting the difficulty of distinguishing this species.

At IoU = 0.5, classification accuracy declined across all species. Correct classifications for *L. buriatica* decreased to 859, while those for *N. pulla* and *A. scoparia* declined to 187 and 22, respectively. The normalized confusion matrix shows increased off-diagonal values. Across all evaluated weed species, overall detection accuracy ranged between approximately 60% and 80%.

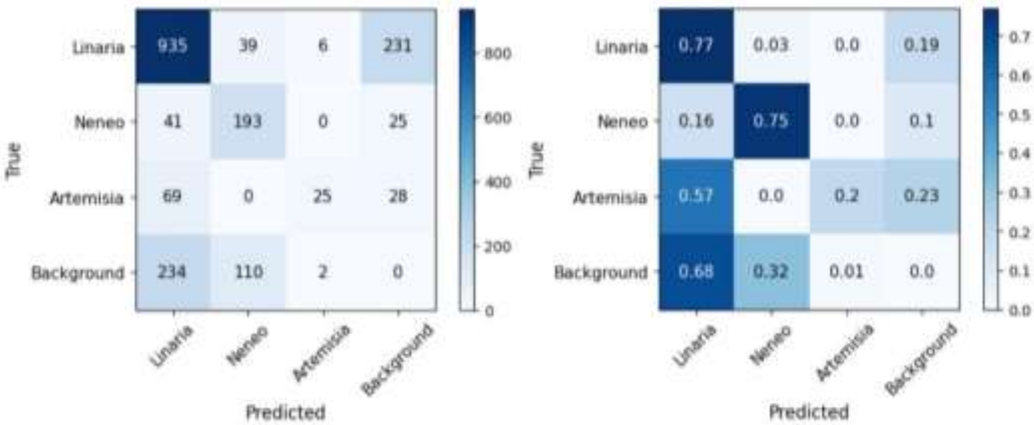


Figure 6. ConfusionMatrix, Model of Mask R-CNN, IoU=0.25.

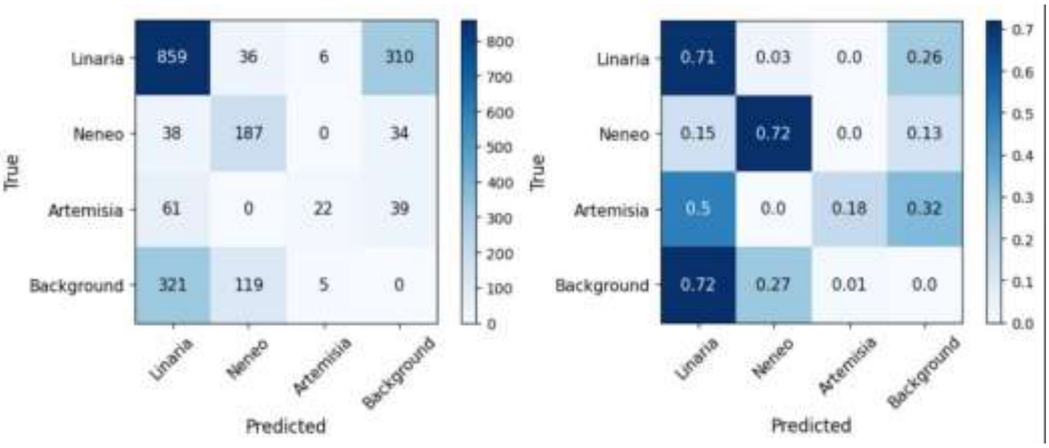


Figure 7. ConfusionMatrix, Model of Mask RCNN, IoU=0.5.

3.3. Model YOLOv8 IoU-Based Detection Performance

For *N. pulla*, the model showed moderate detection performance. Precision reached 0.69, while Recall was lower at 0.48, resulting in an F1-score of 0.57 and an AP of 0.36. Although many detected instances were correctly classified, the relatively high number of false negatives suggests that a substantial portion of true objects was not detected. The weakest performance at IoU = 0.25 was observed for *A. scoparia*. While Precision remained relatively high at 0.49, Recall was limited to 0.37, yielding an F1-score of 0.57 and an AP of 0.36. Only 55 true positives were identified. Across all species, YOLOv8 achieved an overall Precision of 0.65, Recall of 0.53, F1-score of 0.58, and a mean AP (mAP) of 0.59 at IoU = 0.25. When the IoU threshold was increased to 0.5, detection performance decreased for all species. *L. buriatica* remained the best-performing class, but Precision, Recall, and F1-score declined to 0.65, 0.65, and 0.65, respectively, with AP decreasing to 0.51. The number of true positives dropped to 791, accompanied by an increase in both false positives and false negatives. For *N. pulla*, Recall further decreased to 0.45, while Precision remained relatively stable at 0.65, resulting in an F1-score of 0.53 and an AP of 0.33. *A. scoparia* experienced the most pronounced performance degradation, with Precision and Recall falling to 0.42 and 0.31, respectively, and AP decreasing to 0.20. Overall, at IoU = 0.5, YOLOv8 achieved a mean Precision of 0.58, Recall of 0.47, F1-score of 0.36, and mAP of 0.46 (Table 7).

Table 7. Metric Model-Ultralytics YOLOv8.

	Species of weed	**Precision**	**Recall**	**F1 Score**	**AP**	**True Positive**	**False Positive**	**False Negative**
IoU: 0.25	*L. buriatica*	0.76	0.75	0.75	0.67	916	287	274
	N. pulla	0.69	0.48	0.57	0.36	124	55	135
	A. scoparia	0.49	0.37	0.42	0.30	54	37	67
	All	0.49	0.37	0.58	0.59	1094	379	476
IoU: 0.5	*L. buriatica*	0.65	0.65	0.65	0.51	791	412	420
	N. pulla	0.65	0.45	0.53	0.33	117	62	142
	A. scoparia	0.42	0.31	0.36	0.20	42	49	80
	All	0.42	0.31	0.51	0.46	950	523	642

Semantic segmentation or pixel-wise classification, of RGB images acquired at the ground level, into vegetation and background is a critical step in the estimation of several canopy traits [41]. In contrast, *Neneo pulla* and *Artemisia scoparia* showed notably lower Recall values, particularly at stricter IoU thresholds. These species have a smaller physical size, fragmented canopy structure, and spectral similarity to surrounding crops and soil background, which leads to a higher proportion of missed detections, while the biological, ecological, and impact-related characteristics of weeds determine their spatial distribution, spectral appearance, and detectability in UAV imagery, thereby influencing detection accuracy and site-specific weed management strategies discussed in the following sections [42].

3.4. Confusion Matrix Results of Ultralytics YOLOv8

To further examine classification behavior, confusion matrices were analyzed at IoU thresholds of 0.25 and 0.5 using both normalized proportions and raw detection counts. At IoU = 0.25, *L. buriatica* exhibited the highest classification accuracy, with approximately 77% of samples *L. buriatica* correctly predicted. A total of 916 instances were classified correctly, confirming the strong performance observed in the IoU-based metrics. Some misclassification with *N. pulla* and *A. scoparia* was observed, but overall class separability remained high. For *N. pulla*, classification performance was moderate, with 124 true positives. Misclassification between *L. buriatica* and *A. scoparia* classes occurred frequently. For *A. scoparia* 54 instances correctly classified. But a substantial number of samples were misclassified into other categories. At IoU = 0.5, classification accuracy declined across all species. Correct classifications for *L. buriatica* decreased to 791, while those for *N. pulla* and *A. scoparia* declined to 117 and 42, respectively. This trend reflects the combined challenges of stricter localization requirements and inter-class visual similarity (Figures 8 and 9).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0402.v1

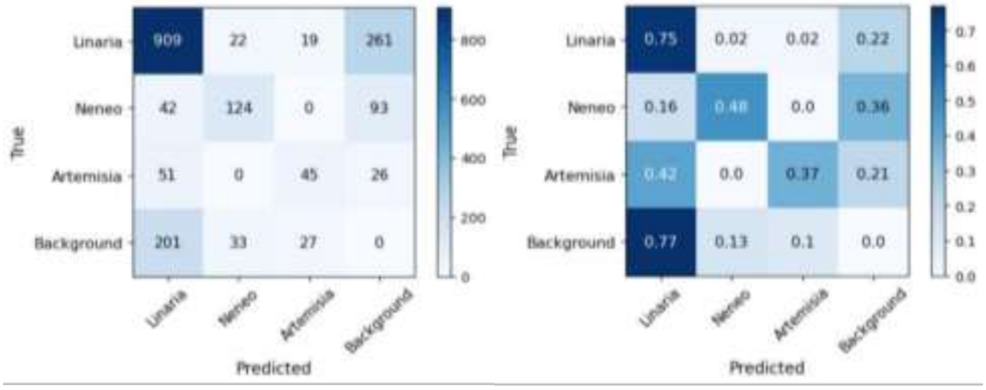


Figure 8. Confusion matrix—Model of YOLOv8, IoU = 0.25.

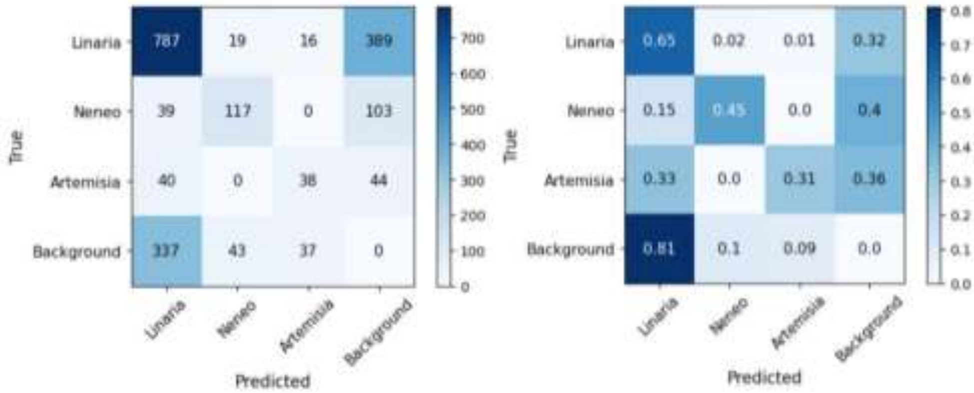


Figure 9. Confusion matrix—Model of YOLOv8, IoU = 0.5.

4. Discussions

4.1. IoU-Based Metrics of Mask R-CNN

The IoU-based results demonstrate that Mask R-CNN performs effectively for weed species with larger size and well-defined morphological features. *L. buriatica* consistently achieved the highest performance at both IoU thresholds, reflecting its clear structural boundaries and higher representation in the dataset. For *N. pulla*, the model exhibited strong Recall but comparatively lower Precision, indicating a tendency to generate a large number of candidate regions. While this behavior increases the likelihood of detecting true instances, it also leads to increased false positives, suggesting limitations in classification refinement. The weakest performance was consistently observed for *A. scoparia*, where recall values remained below 0.21 across both IoU thresholds. Despite relatively high Precision, the low Recall indicates that Mask R-CNN often failed to generate region proposals for this species. This limitation is likely associated with its small size, fragmented structure, and low visual contrast, which reduce the effectiveness of region proposal generation. The decline in performance when increasing the IoU threshold from 0.25 to 0.5 across all species highlights Mask R-CNN's sensitivity to mask boundary accuracy. While the model maintains relatively stable localization for medium-to-large weeds, precise segmentation of fine and irregular plant structures remains challenging under real field conditions. Mask R-CNN demonstrates that the additional mask output differs from the class and bounding box outputs, requiring a more precise extraction of the object's spatial structure and localization [16].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 April 2026 doi:10.20944/preprints202604.0402.v1

4.2. Confusion Matrix Analysis

The confusion matrix analysis further supports the trends observed in the IoU-based metrics. The strong diagonal dominance for *L. buriatica* confirms that species with consistent and well-defined visual features are reliably classified. In contrast, the increased confusion between *L. buriatica* and *N. pulla* suggests partial overlap in spectral and morphological characteristics. The pronounced misclassification of *A. scoparia* across both IoU thresholds indicates poor class separability for this species. This issue becomes more severe at IoU = 0.5, where accurate detection and segmentation are required in addition to correct classification. These findings suggest that, although Mask R-CNN is well suited for detecting larger and visually distinct weeds, its performance is constrained when applied to small, irregular, or visually ambiguous species. At lower IoU thresholds, both Mask R-CNN and YOLOv8 achieved higher Recall values, indicating that the models were generally effective at identifying the majority of weed instances present in the field.

Under minor overlap constraints, both architectures were able to capture coarse object location. However, as the IoU threshold increased, a systematic decline in Recall and average Precision was observed. This reduction reflects the increasing difficulty of achieving precise boundary delineation, particularly in areas characterized by dense vegetation, overlapping plants, and variable illumination conditions. This behavior is consistent with findings reported in previous UAV-based weed detection studies, which highlight that segmentation accuracy is often constrained by fine-scale plant morphology, partial occlusion, and low contrast between target weeds and surrounding crops or soil. Consequently, while higher IoU criteria provide a more stringent and meaningful evaluation of segmentation quality, they also expose inherent limitations of current deep learning–based detection models when applied to complex real-world agricultural settings.

4.3. IoU-Based Metrics of YOLOv8

The detection performance of the YOLOv8 model was first evaluated using standard IoU-based metrics at thresholds of 0.25 and 0.5. At an IoU threshold of 0.25, *L. buriatica* achieved the strongest detection performance among all species. The model obtained a Precision of 0.76, Recall of 0.77, and an F1-score of 0.76, with an AP of 0.67. A total of 916 true positives were detected, indicating that YOLOv8 was highly effective in both detecting and localizing this species under minor overlap constraints. The IoU-based results revealed clear performance differences across weed species, largely driven by variations in plant size, morphology, and visual distinctiveness. *L. buriatica* consistently achieved the highest performance at both IoU thresholds, which can be attributed to its relatively large size, well-defined structure, and strong contrast with the background. These characteristics are well aligned with YOLOv8's single-stage detection architecture, facilitating effective feature extraction and bounding box regression.

In contrast, *N. pulla* demonstrated stable Precision but consistently lower Recall, indicating that while detected instances were generally accurate, many true objects were missed. This behavior is likely related to the visual similarity of this species to surrounding vegetation, which reduces the discriminative power of extracted features. The weakest IoU-based performance was observed for *A. scoparia*, particularly at the higher IoU threshold. The combination of low Recall and declining AP suggests that YOLOv8 struggles to reliably detect this species, especially when precise localization is required. Its small size, fragmented growth pattern, and low contrast against the background likely limit the effectiveness of anchor-based detection. The overall decline in performance when increasing the IoU threshold from 0.25 to 0.5 highlights YOLOv8 sensitivity to detection accuracy. While the model performs well under minor overlap conditions, achieving precise spatial alignment remains challenging in complex field environments.

4.4. Confusion Matrix Analysis of YOLOv8

The confusion matrix analysis provides deeper insight into YOLOv8 classification behavior. The strong diagonal dominance observed for *L. buraitica* confirms that species with clear and consistent

 doi:10.20944/preprints202604.0402.v1

visual features can be reliably distinguished. In contrast, the increased off-diagonal entries for *N. pulla* and *A. scoparia* indicate persistent class mixed-up, particularly under stricter IoU constraints. The pronounced misclassification of *A. scoparia* suggests that YOLOv8 often fails to extract sufficiently discriminative features for small, irregular, and visually ambiguous plants. Overall, the confusion matrix results are consistent with the IoU-based metrics and reinforce the conclusion that YOLOv8 is well suited for detecting medium-to-large weed species with well-defined morphological characteristics but remains limited when applied to small or visually complex weeds. Improving performance for such species may require class-balanced training data, enhanced multi-scale feature representation, and targeted data augmentation strategies.

4.5. Comparative Analysis of Mask R-CNN and YOLOv8 Performance

The evaluation metrics provide clear numerical evidence of the complementary performance characteristics of YOLOv8 and Mask R-CNN. Both models classified and detected crops, single weeds, and weed patches, while also segmenting each detected object to provide detailed information about their boundaries and regions [43]. At an IoU threshold of 0.25, YOLOv8 achieved a slightly higher overall F1-score (0.72) than Mask R-CNN (0.71), together with higher Precision (0.74 versus 0.69), indicating more stable classification behavior and reduced false-positive rates. This result supports the notion that YOLO models provide a better balance between speed and accuracy for real-time applications in object detection [44]. Confusion matrix analysis (Figures 8 and 9) further supports this observation. YOLOv8 exhibited stronger diagonal dominance for *L. buriatica*, with 916 correctly classified instances compared to 938 for Mask R-CNN, while generating substantially fewer false positives (287 versus 341). In contrast, Mask R-CNN demonstrated a Recall-oriented detection behavior for *N. pulla*. At IoU = 0.25, Mask R-CNN achieved a markedly higher Recall (0.74) than YOLOv8 (0.47), confirming its effectiveness in identifying a larger proportion of true instances.

However, the lower accuracy in detecting the other classes indicates that additional refinement could improve the detection performance [45]. For *A. scoparia*, both models exhibited limited detection capability, reflecting the inherent difficulty of detecting small, fragmented, and visually ambiguous weeds in UAV-based RGB imagery. The model achieved an overall detection accuracy of approximately 70–75% at IoU = 0.25 while maintaining lower false-positive rates. Mask R-CNN, although computationally more demanding, demonstrated stronger Recall and more detailed instance segmentation under relaxed IoU constraints. At a stricter IoU threshold of 0.5, Mask R-CNN maintained a higher mean Average Precision (mAP = 0.54) compared to YOLOv8 (mAP = 0.46), indicating greater robustness under precise localization requirements [44]. Together, these results quantitatively justify the inclusion of both architectures, highlighting the trade-off between segmentation accuracy and computational efficiency. This trend reflects the expected sensitivity of two-stage instance segmentation models to stricter localization criteria, while still maintaining comparatively strong detection completeness. A higher Recall implies that fewer weed instances are missed, and improved mask IoU reflects more accurate pixel-level delineation of weed boundaries. Together, these characteristics enable more reliable weed coverage estimation and spatial distribution mapping, which are essential for site-specific herbicide management and precision agriculture applications [46].

The substantially reduced inference time allows rapid processing of high-resolution imagery, which is critical for time-sensitive decision-making during the growing season. Previous studies have also shown that modern single-stage architectures such as YOLOv8 can effectively replace traditional two-stage models in instance segmentation tasks where rapid processing is prioritized, without a substantial loss in detection reliability [47]. Nevertheless, YOLOv8 exhibited comparatively lower mask IoU values, indicating less precise boundary segmentation. This limitation can lead to systematic underestimation of weed-covered areas, particularly for species with fine leaves, overlapping structures, or complex morphology. These results underscore a fundamental trade-off between segmentation accuracy and computational efficiency: while Mask R-CNN is more suitable for detailed weed mapping and accurate coverage estimation, YOLOv8 offers clear advantages in

 doi:10.20944/preprints202604.0402.v1

scenarios where speed, scalability, and real-time deployment are prioritized over precise boundary delineation. The superior Recall and mask IoU achieved by Mask R-CNN indicate its strong capability to capture small, fragmented, and irregularly shaped weeds, which are typical under real-field agricultural conditions [46].

A higher Recall implies that fewer weed instances are missed during detection, while an improved mask IoU reflects more accurate pixel-level delineation of weed boundaries. Together, these characteristics enable more reliable weed coverage estimation and spatial distribution mapping, which are critical for precision agriculture applications such as site-specific herbicide management. These characteristics make YOLOv8 more suitable for large-area monitoring and near–real-time decision-making, where rapid processing is prioritized over fine-scale boundary accuracy. The relatively low mAP score juxtaposed with the high mAP and F1 score highlights an area for further research and development, particularly in enhancing the detection accuracy of smaller weed instances [48]. This difference in processing time showed suitability of the YOLOv8 for both single and multi-object instance segmentation for real-time applications, particularly in agricultural settings where computational resources may be scarce [28].

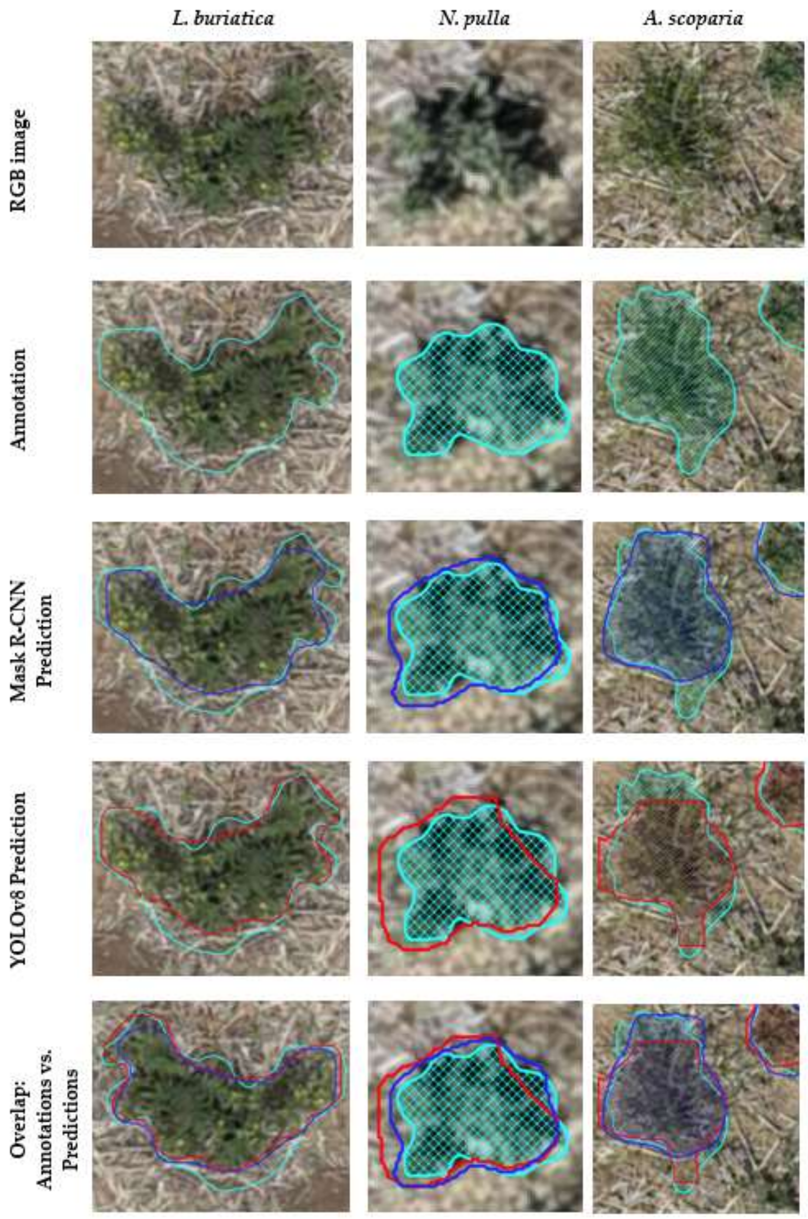


Figure 10. The results of the model-generated visualization were compared with those manually annotated.

When comparing the metrics at IoU = 0.25 and IoU = 0.5, the results indicate that IoU = 0.25 achieved superior performance across all evaluation metrics. This outcome may be influenced by factors such as plant characteristics, morphological features, the composition of the dataset, and the parameters of the applied model (Table 8).

Table 8. IoU evaluation metrics model of Mask-RCNN and YOLO v8.

Metrics	Precision				Recall				F1 Score				AP			
Models	Mask R-CNN		YOLOv8		Mask R-CNN		YOLOv8		Mask R-CNN		YOLOv8		Mask R-CNN		YOLOv8	
IoU	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50
L. buriatica	0.73	0.67	0.76	0.65	0.77	0.71	0.75	0.65	0.75	0.69	0.75	0.65	0.69	0.59	0.67	0.51
N. pulla	0.56	0.55	0.69	0.65	0.75	0.72	0.48	0.45	0.64	0.62	0.57	0.53	0.56	0.53	0.36	0.33
A. scoparia	0.76	0.67	0.49	0.42	0.20	0.18	0.37	0.31	0.32	0.28	0.42	0.36	0.17	0.13	0.30	0.20

4.6. Species-Specific Model Performance

At the core of effective weed detection are the curation of a large-scale, precisely labelled weed dataset and whereby the development of supervised learning models [49]. Model performance varied substantially across weed species, highlighting the influence of species-specific morphology, visual separability, and data balance on detection accuracy [50]. Among the evaluated species, *L. buriatica* consistently achieved the highest Precision, Recall, and F1-scores across IoU thresholds. At IoU = 0.25, *L. buriatica* reached a Recall of 0.78 and an F1-score of 0.75, indicating effective detection and segmentation for visually distinctive and structurally well-defined species. In contrast, *N. pulla* exhibited moderate performance, characterized by relatively stable Precision but lower Recall and F1-scores, suggesting a higher rate of omission errors. *A. scoparia* showed the poorest performance, with Recall values as low as 0.20 at IoU = 0.25 and 0.18 at IoU = 0.5, accompanied by low AP values, indicating significant detection challenges for small and fragmented weed species. These findings align with recent UAV-based weed detection studies showing that DL models achieve higher performance for easily distinguishable species, while smaller weed were difficult to detect [30]. Systematic analyses of DL-driven weed detection also report the prevalence of species-dependent variability and emphasize the need for class-balanced and morphology-aware sample preparation [51].

4.7. Theoretical Comparison of Mask R-CNN and YOLOv8

The observed performance differences between Mask R-CNN and YOLOv8 can be attributed to the fundamental differences in detection paradigms. On the one hand Mask R-CNN two-stage pipeline enables region proposal followed by refinement, resulting in higher Recall and more accurate spatial alignment instances [16]. On the other hand, YOLOv8 single-stage architecture prioritizes high-confidence detections and computational efficiency, which results in faster inference but increased omission errors for visually ambiguous samples [51]. Comparative DL literature also shows that single-stage methods, including YOLOv8, often achieve higher Precision and competitive Recall under optimal conditions but may underperform two-stage methods for boundary-sensitive tasks [52].

4.8. Confidence-Based Trade-Off and Curve Interpretations

Confidence-based performance further illustrated the trade-off between Precision and Recall. YOLOv8 maintained consistently high Precision across increasing confidence thresholds, indicating effective suppression of false positives. However, Recall declined sharply at higher confidence levels, particularly for *N. pulla* and *A. scoparia*. The optimal F1-score across all classes was observed at a confidence threshold of approximately 0.45, suggesting that highly conservative thresholds may exacerbate omission errors. Similar confidence–Recall trade-offs have been documented in other

 doi:10.20944/preprints202604.0402.v1

agricultural object detection contexts, where high-confidence detectors favor Precision at the cost of Recall, necessitating careful threshold tuning for field-scale applications [30]. Analysis of Precision–Recall and F1-score curves revealed a clear trade-off between false positives and false negatives as confidence thresholds varied. The reduction in average precision observed for increased IoU thresholds indicates limitations in boundary delineation accuracy, especially for weeds with thin or irregular structures. This effect was most pronounced for *A. scoparia*, whose fragmented morphology poses inherent challenges for mask prediction using RGB imagery alone [53].

4.9. Implications for Precision Weed Management

From an applied agriculture perspective, these findings have direct implications for site-specific weed management. The strong performance of YOLOv8 for visually distinctive and dominant weed species suggests its suitability for rapid field-scale scouting and early warning systems, enabling farmers to identify infestation hotspots efficiently [54,55]. In contrast, Mask R-CNN's higher Recall and superior mask IoU enable more accurate weed boundary delineation and coverage estimation, which are critical for variable-rate herbicide application and decision-support systems [54]. Selecting an appropriate model based on operational objectives prioritizing speed and scalability versus spatial accuracy can substantially enhance the effectiveness and economic efficiency of precision weed management strategies [51].

4.10. Limitations And Future Directions

Despite the promising results, several limitations remain. Class imbalance significantly affected detection accuracy, with minority weed species consistently exhibiting lower Recall and IoU, a well-documented challenge in deep learning–based precision agriculture applications [51]. Future work should investigate class-balanced training strategies, advanced data augmentation, and the integration of multispectral or hyperspectral imagery to improve feature discrimination [53]. The higher Recall achieved by Mask R-CNN for *N. pulla* is likely attributable to its two-stage detection paradigm, which emphasizes exhaustive region proposal generation. This design reduces omission errors for visually ambiguous weeds but increases false positives, as also reflected in the detection counts reported (Table 9).

Table 9. Detection calculation of weed.

Species of weed	Annotation	Mask R-CNN	Difference+/-	Normalized +/-, %	YOLOv8	Difference +/-	Normalized +/-, %
L. buriatica	1211	1279	68	5.61	1203	-8	-0.66
N. pulla	259	342	83	32.04	179	-80	-30.8
A. scoparia	122	33	89	72.95	91	-31	-25.4
All	1592	1654	62	3.89	1473	-119	-7.47

5. Conclusions

This study evaluated the performance of Mask R-CNN and Ultralytics YOLOv8 for UAV-based weed detection and instance segmentation under real-field agricultural conditions using multiple quantitative metrics at IoU thresholds of 0.25 and 0.5. The results revealed a consistent trade-off between detection completeness, segmentation accuracy, and computational efficiency across the two architectures.

Mask R-CNN demonstrated higher Recall, mAP, and mask IoU values, achieving an overall Recall of 0.77 and F1-score of 0.75 at IoU = 0.25, and maintaining robust performance at IoU = 0.5 (Recall = 0.72, F1-score = 0.69). These results indicate strong capability in minimizing omission errors and accurately delineating weed boundaries, which is essential for reliable weed coverage estimation and spatial analysis. In contrast, YOLOv8 achieved higher Precision (0.76 at IoU = 0.25) and lower false-positive rates, but exhibited lower Recall, particularly under stricter localization criteria, leading to underestimation of weed instances in detection count analysis.

Species-level evaluation showed that *L. buriatica* was consistently detected with high accuracy, whereas *N. pulla* and especially *A. scoparia* posed greater challenges due to morphological complexity and limited visualization. Overall, the findings indicate that Mask R-CNN is more suitable for applications requiring detailed weed mapping and accurate coverage estimation, while YOLOv8 offers an efficient solution for real-time UAV-based monitoring and large-area field scouting. Future research should focus on improving detection of more weed species in agricultural fields of Mongolia.

Author Contributions: Conceptualization, S.J., M.L.L.C., M.I. and Y.D.; methodology, S.J., M.L.L.C. and M.I.; software, S.J., M.S.I., Y.D., L.S., K.S. and V.B; validation S.J., M.S.I.; formal analysis, S.J., M.S.I., Y.D. and T.L.; investigation, S.J.; M.S.I., M.L.L.C. and Y.D.; resources, M.I., M.L.L.C. and B.Ch.; data curation, S.J., J.Y., N.A., S.T., T.L. and F.N; writing-original draft preparation, S.J.; writing-review and editing, M.L.L.C., B.B. and T.L.; visualization, M.L.L.C., Y.D., supervision, M.L.L.C., Y.D., T.L., project administration, M.L.L.C.; funding acquisition, M.L.L.C., B.Ch.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the "Engineering and Technological Higher Education" Project under the Medium-Term Program of the Strategic Partnership between Mongolia and Japan, and by the project on the "Drone application in precision agriculture".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors do not have permission to share data.

Acknowledgments: This study aims to support the advancement of methodologies for analyzing weed distribution in future research. In addition, it plays a crucial role in facilitating the adoption of advanced weed identification approaches based on intelligent technologies in Mongolia. The authors would like to express their sincere gratitude to B. Ariunsanaa, a student at the Mongolian University of Life Sciences, for valuable assistance and support during the field research.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
RGB	Red, Green, Blue
DL	Deep Learning
R-CNN	Convolutional Neural Networks
YOLO	You Only Looks Once
U-Net	U-Network
RoI	Region of Interest
IoU	Intersection over Union
SSD	Single Shot Detector
GPU	Graphics Processing Unit
AP	Average Precision
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative
GPS	Global Positioning System
GIS	Geographic Information System

GDP	Gross Domestic Product
FAO	Food and Agriculture Organization

References

1. Radoglou-Grammatikis, P.; Sarigiannidis, P.; Lagkas, T.; Moscholios, I. A Compilation of UAV Applications for Precision Agriculture. *Comput. Networks* **2020**, *172*, doi:10.1016/j.comnet.2020.107148.
2. Protein and Amino Acid Requirements in Human Nutrition: Report of a Joint WHO/FAO/UNU Expert Consultation; World Health Organization, 2007; ISBN 9789241209359.
3. Gyeltshen, K.; Sharma, S. Integrated Plant Nutrition System Modules for Major Crops and Cropping Systems in South Asia; 2019; ISBN 9789843485922.
4. Ministry of Food, A. and L.I. of M. Crop Plantation Progress as of 3 June 2025.
5. Ying, B.; Xu, Y.; Zhang, S.; Shi, Y.; Liu, L. Weed Detection in Images of Carrot Fields Based on Improved YOLO V4. *Trait. du Signal* **2021**, *38*, 341–348, doi:10.18280/TS.380211.
6. Sivakumar, A.N.V.; Li, J.; Scott, S.; Psota, E.; Jhala, A.J.; Luck, J.D.; Shi, Y. Comparison of Object Detection and Patch-Based Classification Deep Learning Models on Mid-to Late-Season Weed Detection in UAV Imagery. *Remote Sens.* **2020**, *12*, doi:10.3390/rs12132136.
7. И.Оттонбаатар; Т.Аззаяа Сэлгээний Талбайн Хог Ургамлын Судалгаа. *Mong. J. Agric. Sci.* **2014**, *10*, 127–132, doi:DOI:10.5564/mjas.v10i1.310.
8. Daraghmi, Y.A.; Naser, W.; Daraghmi, E.Y.; Fouchal, H. Drone-Assisted Plant Stress Detection Using Deep Learning: A Comparative Study of YOLOv8, RetinaNet, and Faster R-CNN. *AgriEngineering* **2025**, *7*, doi:10.3390/agriengineering7080257.
9. Maddikunta, P.K.R.; Hakak, S.; Alazab, M.; Bhattacharya, S.; Gadekallu, T.R.; Khan, W.Z.; Pham, Q.-V. Unmanned Aerial Vehicles in Smart Agriculture: Applications, Requirements and Challenges. **2020**, doi:10.1109/JSEN.2021.3049471.
10. Monteiro, A.; Santos, S. Sustainable Approach to Weed Management: The Role of Precision Weed Management. *Agronomy* **2022**, *12*, doi:10.3390/agronomy12010118
11. Saini, P.; Nagesh, D.S. A Review of Deep Learning Applications in Weed Detection: UAV and Robotic Approaches for Precision Agriculture. *Eur. J. Agron.* **2025**, *168*, doi:10.1016/j.eja.2025.127652
12. Zhang, Q.; Liu, Y.; Gong, C.; Chen, Y.; Yu, H. Applications of Deep Learning for Dense Scenes Analysis in Agriculture: A Review. *Sensors (Switzerland)* **2020**, *20*, doi:10.3390/s20051520
13. A., A.; J., A.; O., A.; O., A.; E., M.; A., A. A Convolutional Neural Network Model for Crop Disease Detection System. *Br. J. Comput. Netw. Inf. Technol.* **2024**, *7*, 94–102, doi:10.52589/bjcnit-z1blvyo8.
14. Patrício, D.I.; Rieder, R. Computer Vision and Artificial Intelligence in Precision Agriculture for Grain Crops: A Systematic Review. *Comput. Electron. Agric.* **2018**, *153*, 69–81, doi:10.1016/j.compag.2018.08.001
15. Zhang, J.; Yu, F.; Zhang, Q.; Wang, M.; Yu, J.; Tan, Y. Advancements of UAV and Deep Learning Technologies for Weed Management in Farmland. *Agronomy* **2024**, *14*, doi:10.3390/agronomy14030494.
16. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. **2018**, doi:10.1109/TPAMI.2018.2844175
17. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. *You Only Look Once: Unified, Real-Time Object Detection*, **2016**, doi:10.1109/CVPR.2016.91
18. Sapkota, R.; Ahmed, D.; Karkee, M. Comparing YOLOv8 and Mask R-CNN for Instance Segmentation in Complex Orchard Environments. *Artif. Intell. Agric.* **2024**, *13*, 84–99, doi:10.1016/j.aiia.2024.07.001.
19. Jiao, L.; Zhao, J. A Survey on the New Generation of Deep Learning in Image Processing. *IEEE Access* **2019**, *7*, 172231–172263, doi:10.1109/ACCESS.2019.2956508.
20. Jamsran. Ts; Ulziikhutag. N, S.C. *Улаанбаатар Орчмын Ургамал Таних Бичиг*; Жамран. Ц, Өлзийхутаг. Н.С.Ч., Ed.; https://paleo.ac.mn/post/17146: Улаанбаатар, 1972;
21. Өлзийхутаг. Н; Санчир. Ч; Ургамал.М Монгол Орны Ургамлын Аймаг, Ангилалзүйн Судалгаа Сүүлийн 35 Жилд. *Ботаникийн хүрээлэнгийн эрдэм шинжилгээний бүтээл* **2009**, *21*, 10–28.
22. Ogtonbaatar, I.; Azzaya, T.; Otgonsuren, M. Distribution of Weed Species in the Agricultural Regions Differential by Climate. *Mong. J. Agric. Sci.* **2019**, *26*, 135–140, doi:10.5564/mjas.v26i01.1208.

23. Deluxni, N.; Sudhakaran, P.; Alroobaea, R.; Almotiri, J.; Yousef, A. Adaptive Lightweight Mask R-CNN Model for Underwater Debris Instance Segmentation and Classification towards Sustainable Marine Waste Management. *Sci. Rep.* **2026**, doi:10.1038/s41598-026-44542-0.
24. Jahan, M.K.; Bhuiyan, F.I.; Amin, A.; Mridha, M.F.; Safran, M.; Alfarhood, S.; Che, D. Enhancing the YOLOv8 Model for Realtime Object Detection to Ensure Online Platform Safety. *Sci. Rep.* **2025**, *15*, doi:10.1038/s41598-025-08413-4.
25. Iqra; Giri, K.J. SO-YOLOv8: A Novel Deep Learning-Based Approach for Small Object Detection with YOLO beyond COCO. *Expert Syst. Appl.* **2025**, *280*, doi:10.1016/j.eswa.2025.127447.
26. Rana, A.; Vaidya, P. YOLO-Based Deep Learning Framework for Real-Time Multi-Class Plant Health Monitoring in Precision Agriculture. *Sci. Rep.* **2026**, *16*, doi:10.1038/s41598-025-29132-w.
27. Lu, C.; Gehring, K.; Kopfinger, S.; Bernhardt, H.; Beck, M.; Walther, S.; Ebertseder, T.; Minceva, M.; Hu, Y.; Yu, K. Weed Instance Segmentation from UAV Orthomosaic Images Based on Deep Learning. *Smart Agric. Technol.* **2025**, *11*, doi:10.1016/j.atech.2025.100966.
28. Sapkota, R.; Ahmed, D.; Karkee, M. Comparing YOLOv8 and Mask RCNN for Object Segmentation in Complex Orchard Environments, **2024**, 13, doi:10.1016/j.aiia.2024.07.001
29. Li, Y.; Guo, Z.; Shuang, F.; Zhang, M.; Li, X. Key Technologies of Machine Vision for Weeding Robots: A Review and Benchmark. *Comput. Electron. Agric.* **2022**, *196*, doi:10.1016/j.compag.2022.106880.
30. Silva, J.A.O.S.; Siqueira, V.S. de; Mesquita, M.; Vale, L.S.R.; Marques, T. do N.B.; Silva, J.L.B. da; Silva, M.V. da; Lacerda, L.N.; Oliveira-Júnior, J.F. de; Lima, J.L.M.P. de; et al. Deep Learning for Weed Detection and Segmentation in Agricultural Crops Using Images Captured by an Unmanned Aerial Vehicle. *Remote Sens.* **2024**, *16*, doi:10.3390/rs16234394.
31. Sarker, T.T.; Ahmed, K.R.; Islam, T.; Rankrape, C.B.; Gage, K. WeedSense: Multi-Task Learning for Weed Segmentation, Height Estimation, and Growth Stage Classification. **2025**, doi:10.48550/arXiv.2508.14486
32. Ahmad, I.; Shang, F.; Faheem, M. Enhanced YOLOv8 Model for Accurate and Real-Time Remote Sensing Target Detection. *Comput. Stand. Interfaces* **2026**, *96*, doi:10.1016/j.csi.2025.104093.
33. Padilla, R.; Passos, W.L.; Dias, T.L.B.; Netto, S.L.; Da Silva, E.A.B. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electron.* **2021**, *10*, 1–28, doi:10.3390/electronics10030279.
34. Talha, M.M.; Khan, H.U.; Iqbal, S.; Alghobiri, M.; Iqbal, T.; Fayyaz, M. Deep Learning in News Recommender Systems: A Comprehensive Survey, Challenges and Future Trends. *Neurocomputing* **2023**, *562*, doi:10.1016/j.neucom.2023.126881
35. Reveles-Martínez, R.; Gamboa-Rosales, H.; Sánchez-Femat, E.; Saldívar-Pérez, J.; Ibarra-Pérez, T.; Reveles-Gómez, L.C.; Guirette-Barbosa, O.A.; Galván-Tejada, J.I.; Galván-Tejada, C.E.; Luna-García, H.; et al. Benchmarking YOLOv8 to YOLOv11 Architectures for Real-Time Traffic Sign Recognition in Embedded 1:10 Scale Autonomous Vehicles. *Technologies* **2025**, *13*, doi:10.3390/technologies13110531.
36. Komisarek, M.; Pawlicki, M.; D'Antonio, S.; Kozik, R.; Pawlicka, A.; Choras, M. Enhancing Network Security Through Granular Computing: A Clustering-by-Time Approach to NetFlow Traffic Analysis. In Proceedings of the ACM International Conference Proceeding Series; Association for Computing Machinery, July 30 **2024**, doi:10.1145/3664476.3670882
37. Kermal, O., Baris, C.C., Emre.A., Sinan. K., *Localization Recall Precision (LRP): A New Performance Metric for Object Detection*, **2018,** doi:10.1007/978-3-030-01234-2_31
38. Zhu, H.; Wei, H.; Li, B.; Yuan, X.; Kehtarnavaz, N. A Review of Video Object Detection: Datasets, Metrics and Methods. *Appl. Sci.* **2020**, *10*, 1–24, doi:10.3390/app10217834
39. Moazzam, S.I.; Khan, U.S.; Qureshi, W.S.; Tiwana, M.I.; Rashid, N.; Alasmary, W.S.; Iqbal, J.; Hamza, A. A Patch-Image Based Classification Approach for Detection of Weeds in Sugar Beet Crop. *IEEE Access* **2021**, *9*, 121698–121715, doi:10.1109/ACCESS.2021.3109015.
40. Jung, H.K.; Choi, G.S. Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions. *Appl. Sci.* **2022**, *12*, doi:10.3390/app12147255.
41. Madec, S.; Irfan, K.; Velumani, K.; Baret, F.; David, E.; Daubige, G.; Samatan, L.B.; Serouart, M.; Smith, D.; James, C.; et al. VegAnn, Vegetation Annotation of Multi-Crop RGB Images Acquired under Diverse Conditions for Segmentation. *Sci. Data* **2023**, *10*, doi:10.1038/s41597-023-02098-y.

42. Hammad, S.Y.; Kovács, G.P.; Milics, G. Ambrosia Artemisiifolia in Hungary: A Review of Challenges, Impacts, and Precision Agriculture Approaches for Sustainable Site-Specific Weed Management Using UAV Technologies. *AgriEngineering* **2026**, *8*, 30, doi:10.3390/agriengineering8010030.
43. Joy, J.; Kelvin, B.; Howatt, K.; Aderholdt, W.; Khan, M.; Peters, T.; Sun, X. Edge-Deployable Segmentation and Prescription Mapping of Post-Emergence Weeds in Sugar Beet Crops for UAV-Based Precision Spraying. *J. Agric. Food Res.* **2025**, *24*, doi:10.1016/j.jafr.**2025**.102422.
44. Sharma, A.; Kumar, V.; Longchamps, L. Comparative Performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN Models for Detection of Multiple Weed Species. *Smart Agric. Technol.* **2024**, *9*, doi:10.1016/j.atech.**2024**.100648.
45. Karim, M.R.; Reza, M.N.; Ahmed, S.; Lee, K.H.; Sung, J.; Chung, S.O. Detection of Trees and Objects in Apple Orchard from LiDAR Point Cloud Data Using a YOLOv5 Framework. *Electron.* **2025**, *14*, doi:10.3390/electronics14132545.
46. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. **2018**, doi:10.48550/arXiv.1804.02767
47. Thenmozhi, K.; Srinivasulu Reddy, U. Crop Pest Classification Based on Deep Convolutional Neural Network and Transfer Learning. *Comput. Electron. Agric.* **2019**, *164*, doi:10.1016/j.compag.2019.104906.
48. Khan, A.T.; Jensen, S.M.; Khan, A.R. Advancing Precision Agriculture: A Comparative Analysis of YOLOv8 for Multi-Class Weed Detection in Cotton Cultivation. *Artif. Intell. Agric.* **2025**, *15*, 182–191, doi:10.1016/j.aiia.2025.01.013.
49. Dang, F.; Chen, D.; Lu, Y.; Li, Z. YOLOWeeds: A Novel Benchmark of YOLO Object Detectors for Multi-Class Weed Detection in Cotton Production Systems. **2023**, doi:10.5281/zenodo.7535814.
50. Hasan, A.S.M.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G.K. A Survey of Deep Learning Techniques for Weed Detection from Images. **2021**, doi:10.1016/j.compag.2021.106067
51. Sandoval-Pillajo, L.; García-Santillán, I.; Pusdá-Chulde, M.; Giret, A. Weed Detection Based on Deep Learning from UAV Imagery: A Review. *Smart Agric. Technol.* **2025**, *12*, doi:10.1016/j.atech.2025.101147
52. Allmendinger, A.; Saltık, A.O.; Peteinatos, G.G.; Stein, A.; Gerhards, R. Assessing the Capability of YOLO- and Transformer-Based Object Detectors for Real-Time Weed Detection. *Precis. Agric.* **2025**, *26*, doi:10.1007/s11119-025-10246-0.
53. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* 2018, *147*, 70–90.
54. Slaughter, D.C.; Giles, D.K.; Downey, D. Autonomous Robotic Weed Control Systems: A Review. *Comput. Electron. Agric.* **2008**, *61*, 63–78, doi:10.1016/j.compag.2007.05.008.
55. Dinis, L.T.; Bernardo, S.; Matos, C.; Malheiro, A.; Flores, R.; Alves, S.; Costa, C.; Rocha, S.; Correia, C.; Luzio, A.; et al. Overview of Kaolin Outcomes from Vine to Wine: Cerceal White Variety Case Study. *Agronomy* **2020**, *10*, doi:10.3390/agronomy10091422