Autonomous Multi-Model Vision System for Site-Specific Weed Detection and Selective Spraying in Smart Agriculture

Benchmarking YOLOv5 and MobileNet-SSD for Field-Deployable Crop-Weed Discrimination

Sana I. Shaikh [1]* and Mayur Jakhete [2]

[1] Department of Electronics Engineering, School of Engineering & Technology (SOET), Pimpri Chinchwad University, Pune, Maharashtra, India
[2] Department of Electronics Engineering, School of Engineering & Technology (SOET), Pimpri Chinchwad University, Pune, Maharashtra, India

*sana.naikwadi1990@gmail.com, mayur.jakhete@pcu.edu.in
*Corresponding author.

Abstract. The impact of weed infestation is a serious threat to global food security. Weeds can reduce crop yield by as much as 43%. Mechanical, manual, and chemical methods have been widely used to control weeds. However, the rising costs of labor, the development of herbicide resistance in weeds, and restrictions on chemical warfare are significant concerns for modern agriculture. This article introduces an autonomous, multimodal robotic solution that scans its environment to detect and remove weeds. Initially, we used 3,120 images of crops, weeds, and backgrounds extracted from Internet resources and field photos. These images were annotated, oriented, and resized to 416 × 416 pixels. Moreover, we applied flips and increased the contrast and exposure to obtain an extended data set. We trained and compared two popular object detectors, including YOLOv5 and MobileNet-SSD, to find the optimal balance between performance and a compact model for further deployment. The mean Average Precision (mAP) of YOLOv5 at 0.5 was approximately 0.86, while the maximum F1-score was 0.83 at a confidence threshold of around 0.49. MobileNet-SSD as a classifier has demonstrated strong performance with an accuracy of 98%. As a result, we also designed and built a mechanical chassis for the robot. Thus, with the obtained results and the analysis of the comparison between the two approaches, we prioritize the following steps for prototype development and field testing.

Keywords: Precision agriculture, weed detection, YOLOv5, MobileNet-SSD, deep learning, robotic weeding, edge AI, smart farming.

© The Author(s) 2026
P. Bagchi et al. (eds.), *Proceedings of the International Conference on Information Systems Design And Intelligent Applications (ICISDA 2026)*, Advances in Intelligent Systems Research 215,
https://doi.org/10.2991/978-94-6239-821-4_2

1 Introduction

Weeds compete with crops for water, nutrients, light, and space; left unchecked, that competition can reduce yields in some cases by up to 43% and threaten global food security [4], [5]. Arable land per person is shrinking while food demand is expected to rise by about 70% by 2050, so there is less room for inefficient weed control each year.

None of the usual methods avoid these problems. Hand weeding does not scale and puts workers at risk; tillage can compact soil and harm crop roots; and herbicides, though fast-acting, are blunt tools that encourage resistant weeds and contaminate soil and waterways [5], [7]. Farmers are shifting toward sensor- and AI-driven farming, applying treatments only at the right place and time instead of spraying entire fields uniformly [6], [9].

Site-Specific Weed Management (SSWM) puts that idea into practice: treatment intensity is tailored to where weeds actually occur instead of applied across the whole field. Advances in machine vision and deep learning now make SSWM feasible at scale, enabling ground robots and UAVs to detect, localize, and respond to weeds in real time [8], [11], [14]. Most studies stop at reporting detection accuracy and seldom extend their work into a full, field-ready pipeline that includes actuation and on-device inference [13].

This work aims to bridge that gap. We present the design and current progress of an autonomous, multi-model robotic system created for precision weed detection and selective removal. To inform deployment choices on memory-constrained hardware, we evaluate the performance of a one-stage detector (YOLOv5) against an edge-optimized detector (MobileNet-SSD) on a custom dataset. Section 2 reviews related literature; Section 3 details dataset preparation; Section 4 describes training and evaluation of both models; Section 5 provides a direct comparison; Section 6 outlines the robotic platform; and Section 7 summarizes results and outlines remaining work.

2 Related Work

2.1 Conventional and Sensor-Based Weed Control

Manual labour, mechanical tillage, and chemical herbicides still form the backbone of weed control, with each option trading off cost, precision, and environmental impact in different ways. Extensive, repeated herbicide use has led to over 267 documented herbicide-resistant weed species and has contributed to soil and water contamination [5], [7]. That track record is why RGB, multispectral, and hyperspectral imaging have drawn growing interest as non-invasive ways to spot weeds without spraying blind [9].

2.2 Deep Learning for Weed Detection

Convolutional neural networks (CNNs) have become the dominant approach for weed and crop differentiation in field imagery [9]. Within this family, one-stage object detectors from the YOLO line-up (spanning the lightweight YOLOv3-tiny through to

YOLOv5, YOLOv7, and YOLOv8) are favoured for their balance between inference speed and detection accuracy; on corn-field weed-detection benchmarks, YOLOv7 and YOLOv8 have been reported to reach mean average precision figures near 89.9% and 89.4% respectively [9]. Pixel-wise segmentation architectures such as U-Net and Mask R-CNN offer finer-grained crop-weed separation at higher computational cost [9]. Feature-extraction backbones such as AlexNet, VGG, and ResNet dominate classification-centric tasks [9], whereas attention-based transformer architectures (e.g., RT-DETR) are becoming an alternative to CNNs for modeling global spatial relations [9].

2.3 Robotic and UAV-Based Platforms

UAVs can cover large areas of a field in a short time, making them ideal for aerial mapping and spraying, while ground robots can take action at an individual-plant level using machine vision [3], [8], [11], [15]. Platforms such as the NVIDIA Jetson Nano and Xavier NX push inference onto the device itself, cutting latency and removing the dependence on a reliable network connection in the field [6], [12]. Regulation still lags the technology, though: UAV flights generally require a human supervisor, and studies on actually removing weeds robotically are far outnumbered by studies that only detect them [10], [13].

2.4 Open Research Gaps

Three gaps kept surfacing across the literature we reviewed. Publicly available, richly annotated datasets covering more than one or two crop-weed pairs are scarce [7], [10]. Models trained on the few, limited datasets available often overfit and fail to generalize when exposed to real field conditions [5], [7]. Very few studies take a detection model all the way through to real robotic actuation and deployment on an edge device [10], [13]. Our work is aimed squarely at the second and third of these: we benchmark two architectures on a dataset we built ourselves, and we are building a robotic platform to match.

3 Proposed Methodology

3.1 System Overview

When using selective spraying, the initial question to ask is whether the weed is in a frame, and the second is where it is located. This results in a system that aims at detection and localization, rather than whole-image classification. The pipeline begins with the collection of the data sets, then proceeds to the annotation, preprocessing, and augmentation of the data sets. It then trains YOLOv5 and MobileNet-SSD, compares them head-to-head, and exports to a format that can be used at the edge by the manufactured chassis.

3.2 Dataset Collection

We collected 3,120 images of crop, weed, and mixed/background scenes from Google Images, Kaggle, and field photography. It was important that lighting, viewing angle, scale, background clutter, and growth stage varied from image to image so the model would not simply memorize a set of lab-like conditions.

3.3 Annotation

Each image was manually labeled in LabelImg by drawing bounding boxes around the crop and weed regions. Consistency of annotation was important to us from an early stage, since sloppy labeling early on leads to inference errors in the field.



Fig. 1. Sample bounding-box annotations for crop and weed classes.

3.4 Preprocessing

We performed two preprocessing steps before training. First, we removed EXIF rotation metadata so that all images have a consistent upright orientation rather than appearing sideways or rotated. Second, we resized images to 416×416 by fitting within this bound while maintaining aspect ratio and without cropping plant areas, matching the input size expected by YOLOv5 and MobileNet-SSD and keeping feature-map size and GPU memory consumption predictable during training.

3.5 Data Augmentation

Keras' ImageDataGenerator created three augmented copies of each image. We applied horizontal flips with a probability of 50%, reflecting the fact that plants are not consistently oriented in the field. Brightness was modulated by ±33% to represent varying conditions of sun, shade, and cloud cover, while exposure was varied by ±17% to represent the slight over- and under-exposure that occurs in outdoor photography.

4 Model Training and Evaluation

4.1 YOLOv5 Model

We selected YOLOv5 [1] for several reasons, including its ability to run on an edge inference format and a practical speed/accuracy compromise when moving from training on a GPU to running on an onboard robot. The training configuration is given in Table 1.

Table 1. YOLOv5 training configuration.

Parameter	Value
Base weights	yolov5s.pt
Image size	416 × 416
Batch size	32
Epochs	150
Optimizer	SGD
Augmentation (hsv_h, hsv_s, hsv_v)	0.015, 0.7, 0.4
Mosaic	1.0
Flip LR	0.5
Translate	0.1
Scale	0.5

The model trained for 150 epochs. The box, objectness, and classification losses gradually decreased for both the training and validation sets, and the precision and recall curves settled into a plateau for the validation set. These curves leveling off indicates that the model did not oscillate and achieved a proper fit rather than noise fitting.

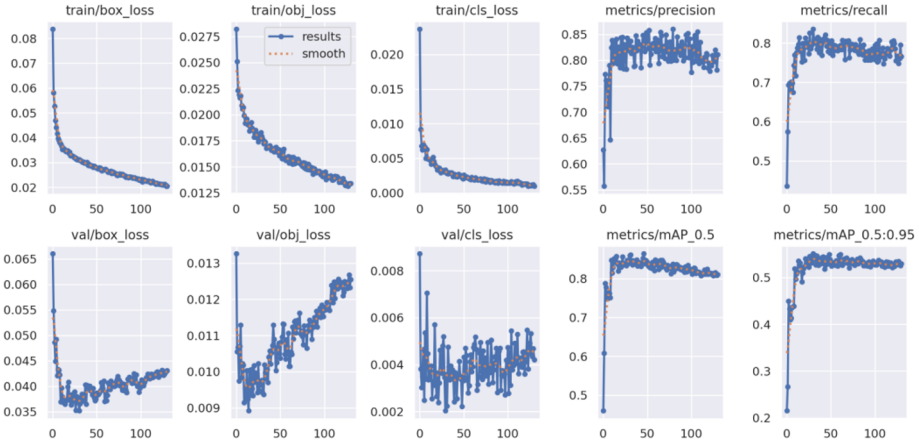


Fig. 2. Training curves for YOLOv5: box, objectness, and classification losses together with precision, recall, mAP@0.5, and mAP@0.5:0.95 across training and validation.

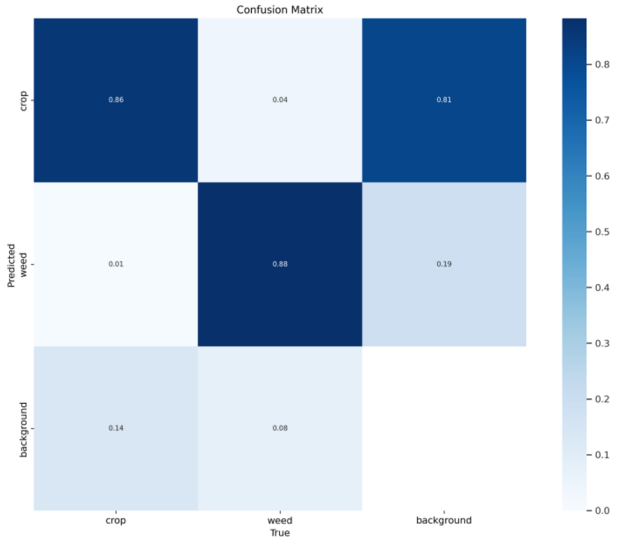


Fig. 3. Confusion matrix for YOLOv5 showing class-wise performance across crop, weed, and background.

Overall, mAP@0.5 was about 0.86. The weed class achieved higher average precision than the crop class, which is desirable since the priority is to detect weeds reliably rather than crops. The best precision–recall balance occurred at roughly a 0.49 confidence threshold, where F1 peaked near 0.83. Raising the threshold increases precision but reduces recall, causing real weeds to be missed; for a spraying system, missing a weed is costlier than triggering an occasional false positive, so we chose a moderate threshold rather than a strict one. The model still has trouble with background and mixed scenes, particularly with small weeds or weeds overlapping the crop canopy, a weakness that appears clearly in the confusion matrix.

4.2 MobileNet-SSD Model

As a lighter alternative, we trained MobileNet-SSD [2], [3] to quantify the accuracy trade-off for faster inference on low-compute hardware. We used transfer learning with a batch size of 32 and trained for 5,000 steps.

Table 2. MobileNet-SSD classification report.

Class	Prec.	Recall	F1	Supp.
Crop	0.98	0.97	0.98	306
Weed	0.97	0.98	0.98	312
Accuracy	—	—	0.98	618
Macro avg	0.98	0.98	0.98	618
Weighted avg	0.98	0.98	0.98	618

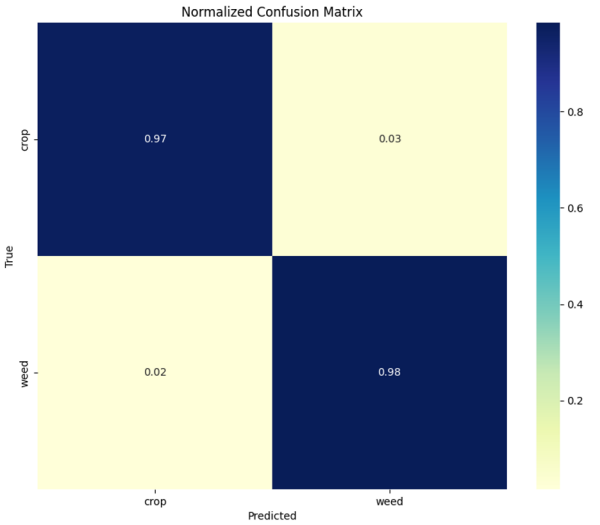


Fig. 4. Normalized confusion matrix for MobileNet-SSD showing class-wise performance between crop and weed.

Classification accuracy was strong, about 97% for crops and 98% for weeds. However, these numbers do not tell the complete story: since MobileNet-SSD is meant to be a detector, its bounding-box quality needs a separate mAP/IoU evaluation, which is still in progress.

5 Comparative Analysis

Table 3 summarizes how the two architectures compare in practice.

Table 3. Comparison of YOLOv5 and MobileNet-SSD.

Aspect	YOLOv5	MobileNet-SSD
Strength	Strong localisation and standard detection metrics (mAP)	Lightweight; high crop/weed classification accuracy

Drawback	Background/mixed scenes cause false positives/negatives	Localisation quality needs separate mAP/IoU evaluation
Best fit	Accurate weed localisation for spraying	Low-compute, fast-inference edge scenarios

YOLOv5 is the better choice when spraying decisions depend on precise weed localisation, since it provides full detection metrics and performs well on small objects. MobileNet-SSD is suitable when speed and low compute are the primary constraints, but it still needs mAP/IoU evaluation before it can be relied on for localisation to the same degree as YOLOv5.

6 Robotic Platform Design and Fabrication

While the detection pipeline was being built, the mechanical side of the project moved forward in parallel: the field platform was modelled in Solid Edge, chassis dimensions were locked down, and the frame itself was laser-cut, then bent and welded into the vehicle base, gear train, and wheel hubs.

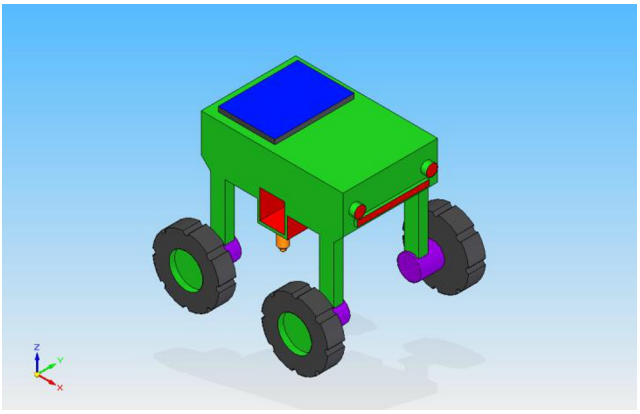


Fig. 5. Proposed 3D assembly of the autonomous weed-detection robotic platform.

A camera-mounted sensing module is built into the design to feed the trained detection models directly; nozzle-triggering logic and the safety interlocks needed for selective herbicide application are next on the list.

7 Conclusion and Future Work

We have described the current state of an autonomous, multi-model robotic system for precision weed detection and selective removal. Starting from a 3,120-image dataset built through collection, annotation, preprocessing, and augmentation, we trained and compared YOLOv5 and MobileNet-SSD: YOLOv5 reached a mAP@0.5 near 0.86

with a peak F1 of 0.83, and MobileNet-SSD reached 98% classification accuracy on crop versus weed, numbers that line up with the accuracy-versus-lightweight-inference trade-off we expected going in. The robot's mechanical chassis was designed and fabricated over the same period.

Several things remain on the roadmap: collecting more manually captured field data to better represent real farms; computing IoU and mAP for MobileNet-SSD so it can be compared against YOLOv5 on equal footing; exporting both models to ONNX/TFLite and clocking their inference speed on edge hardware such as the Jetson Nano; wiring the detection pipeline into the fabricated chassis, complete with camera mounting, nozzle-triggering logic, and safety interlocks; and finally, running the whole selective-spraying system through real-time field trials.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Jocher, G., et al.: YOLOv5. Ultralytics, GitHub repository (2020), https://github.com/ultralytics/yolov5
2. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861 (2017)
3. Liu, W., et al.: SSD: single shot multibox detector. In: Proc. ECCV, pp. 21–37 (2016)
4. Shackleton, C., Ali, R.H., Khan, T.A.: Enhancing rangeland weed detection through convolutional neural networks and transfer learning. Crop Design 3(3), 100060 (2024)
5. Pai, D.G., Kamath, R., Balachandra, M.: Deep learning techniques for weed detection in agricultural environments: a comprehensive review. IEEE Access 12, 113193–113214 (2024)
6. Karim, M.J., Nahiduzzaman, N., Ahsan, M., Haider, J.: Development of an early detection and automatic targeting system for cotton weeds using an improved lightweight YOLOv8 architecture on an edge device. Knowledge-Based Systems 300, 112204 (2024)
7. El Sakka, M., Ivanovici, M., Chaari, L., Mothe, J.: A review of CNN applications in smart agriculture using multimodal data. Sensors 25(2), 472 (2025)
8. Sunil, S.G., et al.: Field-based multispecies weed and crop detection using ground robots and advanced YOLO models: a data and model-centric approach. Smart Agricultural Technology 9, 100538 (2024)
9. Catala-Roman, P., Segura-Garcia, J., Navarro-Camba, E.A., Alcaraz-Calero, J.M., Garcia-Pineda, M.: AI-based autonomous UAV swarm system for weed detection and treatment: enhancing organic orange orchard efficiency with Agriculture 5.0. Internet of Things 28, 101418 (2024)
10. Deng, B., Lu, Y., Xu, J.: Weed database development: an updated survey of public weed datasets and cross-season weed detection adaptation. Ecological Informatics 81, 102546 (2024)
11. Borah, S.S., Anand, A., Fletcher, R., Reddy, K.N., Sundaravadivel, P.S.: Unmanned aerial vehicle solutions for weed detection in precision agriculture. IEEE Access 13, 122457–122471 (2025)

12. Sunil, G.C., Upadhyay, A., Sun, X.: Development of software interface for AI-driven weed control in robotic vehicles, with time-based evaluation in indoor and field settings. Smart Agricultural Technology 9, 100678 (2024)
13. Li, N., Zha, X., Zhang, C., Guo, H., Sun, Z., Wu, X.: Real-time crop recognition in transplanted fields with prominent weed growth: a visual-attention-based approach. Computers and Electronics in Agriculture 203, 107518 (2024)
14. Anam, I., Arafat, N., Hafiz, M.S., Jim, J.R., Kabir, M.M., Mridha, M.F.: A systematic review of UAV and AI integration for targeted disease detection, weed management, and pest control in precision agriculture. Smart Agricultural Technology 9, 100647 (2024)
15. Koshelev, I., Savinov, M., Menshchikov, A., Somov, A.: Drone-aided detection of weeds: transfer learning for embedded image processing. Remote Sensing 16(3), 428 (2024)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

