EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2025, 26(9), 36–44
https://doi.org/10.12912/27197050/208897
ISSN 2719–7050, License CC-BY 4.0

Received: 2025.07.22
Accepted: 2025.08.15
Published: 2025.09.01

Deep learning approach for precise stand counting in Moroccan sugar beet fields

Noura Ouled Sihamman[1*], Assia Ennouni[1], My Abdelouahed Sabri[1], Abdellah Aarab[2]

[1] Department of Computer Science, Faculty of Sciences Dhar el Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco
[2] Department of Physique, Faculty of Sciences Dhar el Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco
* Corresponding author's e-mail: noura.ouledsihamman@usmba.ac.ma

ABSTRACT

Recent advances in artificial intelligence (AI) offer a promising alternative to conventional drone-based remote sensing. In this study, we propose a sustainable stand-enumeration methodology for Moroccan sugar beet fields by integrating high-resolution imagery acquired with unmanned aerial vehicles (UAVs) and a novel deep-learning pipeline. Four convolutional- and transformer-based architectures YOLOv5, Fast R-CNN, YOLOR (YOLOR) and YOLOv7 were trained and evaluated on imagery from the Beni Mellal region using precision, recall, mean average precision (mAP) and plant survival-rate metrics. YOLOv5 achieved the best performance, with 97% precision, 92% recall and 96% mAP, significantly outperforming the other models. These results demonstrate that AI-assisted UAV imagery can deliver highly accurate crop-stand counts, thereby supporting data-driven decision-making and enhancing sustainability in agricultural monitoring.

Keywords: agricultural stand count, Moroccan sugar beet, artificial intelligence, unmanned aerial vehicles, deep learning.

INTRODUCTION

Sugar beet is a vital crop in Morocco's Beni Mellal region, providing key income for local farmers (Woodhill et al., 2022). Accurate stand counting at the seedling and budding stages enables timely decisions such as gap filling, targeted replanting, and optimized irrigation and fertilization (Dhanaraju et al., 2022). that directly impact yield and resource use (Malhi et al., 2021); (Newton, 2021). Conventional manual counts, while straightforward, are labor-intensive, error-prone, and difficult to scale across large fields (Newton, 2022); (Engen et al., 2021).

Recent advances in precision agriculture leverage unmanned aerial vehicles (UAVs) and convolutional neural networks (CNNs) to automate stand estimation (Morim de Lima, 2023).

In crops like maize, canola, rice, and bell pepper, UAV imagery combined with CNN architectures (e.g., U-Net, YOLOv5, YOLOv7) has achieved R^2 values above 0.90 and mean absolute errors below two plants per plot (Pande and Moharir, 2023).

However, these studies focus on agro-climatic contexts outside Morocco and do not address sugar beet's specific morphological changes between early and later growth stages (Arab and Azaitraoui, 2024).

Moreover, existing models are seldom fine-tuned for local field conditions, terrain variability, planting density, canopy contrast, and light conditions, which can degrade performance when applied off the shelf. To date, no study has systematically adapted and evaluated UAV-CNN methods for sugar beet stand counting in Moroccan fields (Volk et al., 2024).

OBJECTIVE AND HYPOTHESES

This study develops a workflow combining high-resolution UAV imaging and locally calibrated CNN models (YOLOv5 and YOLOv7) to count sugar beet stands at the seedling and budding stages in Beni Mellal. We test two hypotheses:

- H1: Locally adapted CNNs will yield $R^2 \geq 0.95$ and MAE ≤ 2 plants.
- H2: Stage-specific models will improve detection robustness by accounting for leaf-size and shape variations.

By filling this methodological gap, our protocol aims to provide a replicable solution for sugar beet and other crops in similar agro-climatic zones.

RELATED WORK

- U-Net for Maize (Vong et al., 2021) Applied U-Net on UAV imagery to segment early maize stands under three tillage systems, achieving R^2 values of 0.95, 0.94, and 0.92.
- YOLOv5 + PointNet for Sorghum (James et al., 2024) Combined YOLOv5 detection on RGB images with a modified PointNet for sorghum panicle point clouds; reported 95.5% accuracy on validation datasets.
- YOLOv5 & U-Net for Canola (Ullah et al., 2024) Used YOLOv5 for plant detection and a lightweight U-Net for row segmentation; achieved 95.6% precision and mIoU of 0.8444.
- YOLOv5 + DeepSORT for Bell Pepper by (Escamilla et al., 2024) Integrated YOLOv5 with DeepSORT for maturity-stage recognition and counting in greenhouses; reached 85.7% accuracy.
- YOLO Variants for Plant Spacing (Wang et al., 2023) Evaluated YOLOv5, YOLOX, and YOLOR on UAV imagery to estimate plant-level spacing variability; YOLOv5 achieved $R^2 = 0.936$ and MAE = 1.958.
- YOLOv4 for Rice (Yeh et al., 2024) Adapted YOLOv4 for rice plant detection on UAV images; reported 97% counting accuracy after activation function tuning.

Table 1 encapsulates the various deep learning architectures and their application in agricultural studies, highlighting the effectiveness of these models in different settings and for various crops. Each study demonstrates significant advancements in plant stand counting and crop management through the use of UAV technology and deep learning models like YOLO and U-Net.

MATERIAL AND METHODOLOGY

Experimental site

The dataset used in this study comes from high-resolution digital images taken using a DJI M300 drone equipped with a ZH20 camera (Li et al., 2022); this was done between March 25, 2022, and October 3, 2022, at various locations, as shown in Figure 1. These images pertain to the cultivation of sugar beets in the Beni Mellal region of Morocco. The images were taken with dimensions of 5184 pixels in width and 3888 pixels in height at altitudes ranging from 10 to 20 meters, allowing for detailed observation.

UAV data collection

This dataset provides a summary of the stand count for sugar beet cultivation. It includes data collected from 6 different parcels. A total of 271 images were taken, containing 56,310 objects (individual sugar beet plants) as shown in

Table 1. Comparison of related works

Study	Model	Crop and context	Metrics
(Vong et al., 2021)	U-Net	Maize (three tillage systems)	R^2 = 0.95, 0.94, 0.92
(James et al., 2024)	YOLOv5 + PointNet	Sorghum panicles (Australia)	Accuracy = 95.5%
(Ullah et al., 2024)	YOLOv5 + U-Net	Canola	Precision = 95.6%, mIoU = 0.8444
Escamilla et (Escamilla et al., 2024)	YOLOv5 + DeepSORT	Bell pepper (greenhouse)	Accuracy = 85.7%
(Wang et al., 2023)	YOLOv5, YOLOX, YOLOR	Plant spacing variability (UAV imagery)	R^2 = 0.936, MAE = 1.958
(Yeh et al., 2024)	YOLOv4	Rice (UAV imagery)	Accuracy = 97%

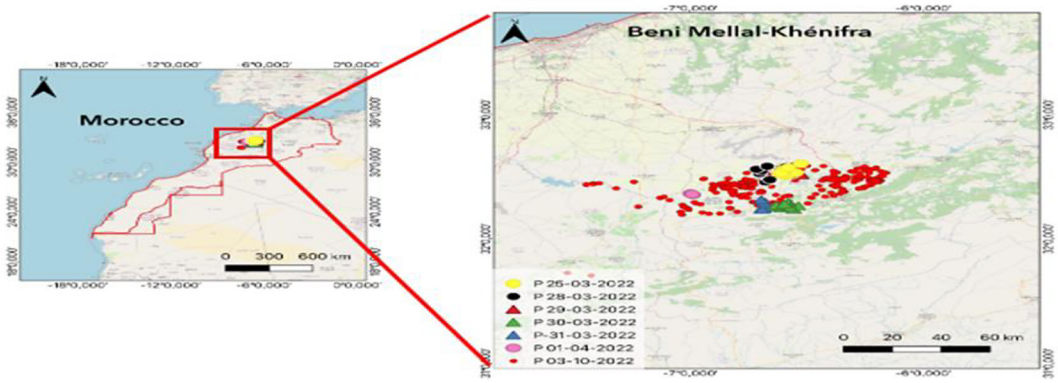

Figure 1. The study area of Beni Mellal-Khenifra in Morocco with the locations of the image samples

Table 2. Dataset used

Parcel	6
Images	271
Number of objects	56310
Number of images	5631

Table 2. On average, each image contains 10 objects, resulting in 5.631 images after normalization to 10 objects per image.

Image processing

To prepare the input layer to meet the requirements of the CNN, all the image preparation steps: resizing, annotating, augmenting, and pruning the data – have been carried out according to the process illustrated in Figure 3. This Figure encompasses all phases of the work, from image acquisition to the training of the deep learning model. As shown in Figure 2.

Images resizing

Due to the large size of the raw photographs, which exceeds the available RAM for processing, each image in the dataset, originally measuring 5184 × 3888 pixels, was resized to 2048 × 1536 pixels. The resizing process utilized an adaptive interpolation method.

Images annotation

To facilitate the annotation process and ensure versatility, we used the COCO format for annotating the images with the VGG Annotator developed by Dutta and Zisserman. This tool allowed us to create precise bounding boxes around each instance of sugar beet, ensuring reliable and consistent annotation. The goal was to accurately delineate each sugar beet plant to facilitate precise and systematic stand counting, as demonstrated in Figure 3, where the annotation was performed by expert agronomists to ensure accuracy and relevance to agricultural practices.

Image annotation for automated analysis was specifically conducted for the 'seedling' and 'budding' stages, as each stage exhibits distinct characteristics such as size, root structure, and nutrient needs, as illustrated in the Figure 4. Additionally, the Table 3 below further explains these differences, which is crucial for ensuring accuracy in identifying plant growth stages and enabling more targeted and effective agronomic interventions.

Data augmentation

Outlines the different data augmentation techniques utilized. Each technique has a unique description, including horizontal and vertical flipping of the image, rotating the image by 90 degrees in various directions, applying grayscale transformation to parts of the images, adjusting hue, saturation, brightness, and exposure, adding blur effects, and introducing noise into the image. These techniques were used to diversify the training data, thereby enhancing the performance of deep learning models in detecting stand counts in sugar beet farms.

Development of the deep learning model

We built our detection pipeline around the YOLOv5 architecture, which streamlines object localization and classification into a coherent, end-to-end workflow. The process unfolds in three primary components: Backbone, Neck, and Head, each optimized for speed and accuracy.

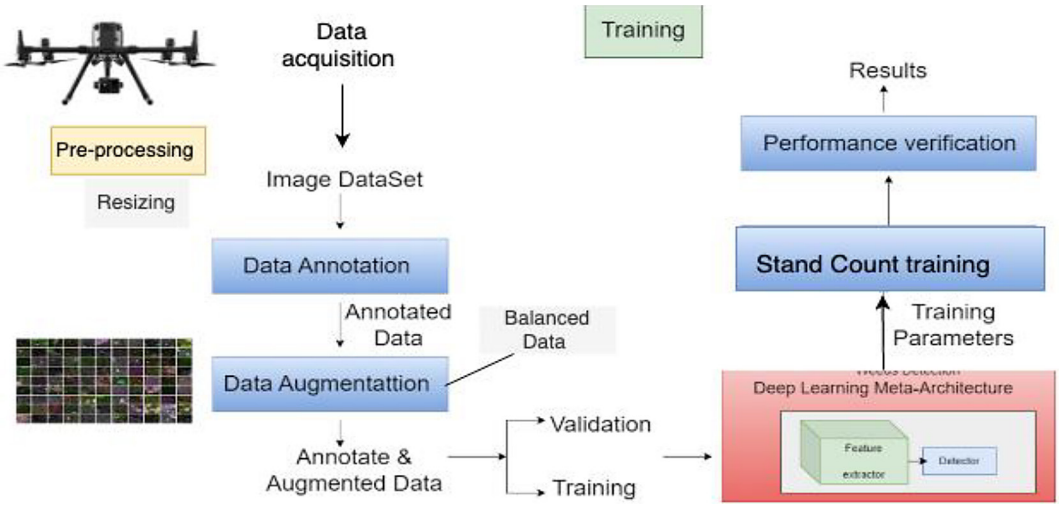

Figure 2. Workflow of automated plant stand detection using deep learning techniques

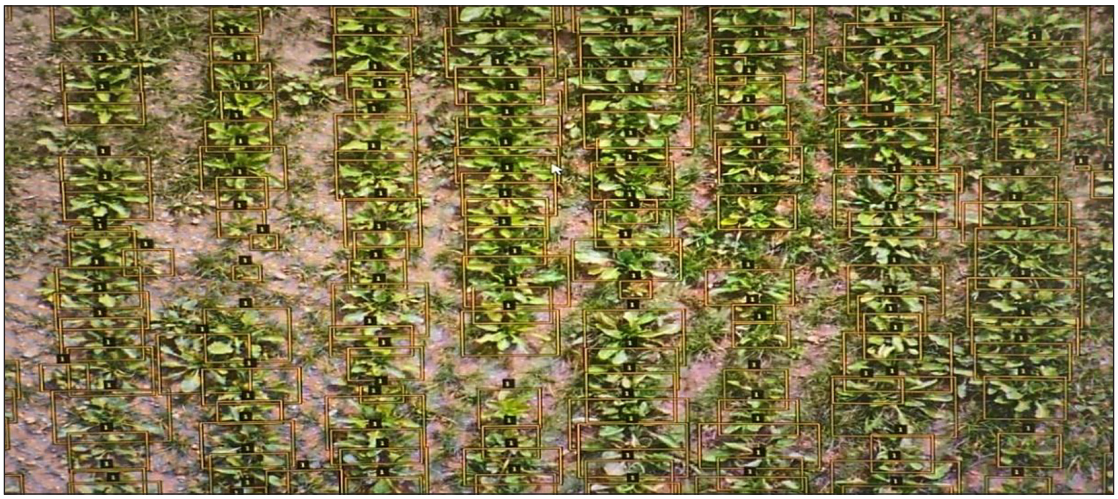

Figure 3. Example of sugar beet stand count annotation using VGG annotator with bounding boxes

- Backbone (feature extraction)
  - Focus layer: The model begins by reorganizing the raw input image into overlapping patches, allowing early layers to capture rich spatial cues (Redmon et al., 2016).
  - Cross stage partial blocks (CSP_L and CSP_X): These modules split and merge feature maps to retain gradient flow while reducing computation, extracting hierarchical representations at multiple scales.
  - Spatial pyramid pooling (SPP): A stack of pooling operations aggregates context from varied receptive fields, making the network robust to objects of different sizes (He et al., 2015).
- Neck (feature aggregation) :
  - Upsampling & concatenation: Feature maps from deep and shallow layers are esampled and merged, preserving both fine details and broader scene context.
  - Squeeze-and-excitation residual units (SE-Res): These blocks recalibrate channel responses, amplifying the most informative features for subsequent detection (Hu et al., 2018).
- Head (prediction) :
  - Prediction layers: The refined feature maps feed into convolutional heads that output bounding-box offsets and class confidences for a set of predefined anchor oxes.
  - Anchor boxes: During training, the network learns several box shapes and aspect ratios suited to the target dataset, enabling precise localization (Redmon and Farhadi, 2018).
  - Grid-based detection: At inference time, the image is divided into a grid of cells;

Table 3. Comparison of seedling and budding stages of sugar beet plants

Characteristics	Seedling stage	Budding stage
General appearance	Young and delicate	More developed and robust
Leaves	Cotyledons visible, first true leaves developing	True leaves well well-developed, forming a rosette
Size	2 to 5 cm (1 to 2 inches)	10 to 20 cm (4 to 8 inches)
Roots	Developing root system, primary taproot, and secondary roots	Extensive and strong root system, thicker taproot, and lateral roots
Buds	No buds	Small buds appearing at leaf axils
Nutrient needs	Requires quickly absorbed, well-balanced nutrients	Needs extra phosphorus for the transition to blooming stage
Color	Green leaves, vibrant color indicating good health	Dark green leaves, thicker stem indicating vigorous growth

Figure 4. Sugar beet growth stages: a – Seedling stage, b – Budding stage

each cell predicts multiple boxes and class scores independently.

- Non-maximum suppression (NMS): Overlapping detections are filtered by retaining only the highest-scoring boxes, which reduces false positives and clutter (Bodla et al., 2017).

Under the hood, YOLOv5 relies on a lightweight yet powerful backbone, often CSPDarknet53 or EfficientNet, augmented with additional convolutional, upsampling, and concatenation layers to ensure thorough coverage of spatial features (Tan and Le, 219). This carefully balanced design delivers real-time performance with high precision, making YOLOv5 an ideal choice for applications ranging from autonomous navigation to precision agriculture. Figure 5 presents a schematic of the full YOLOv5 workflow, illustrating how each component contributes to fast, accurate object detection in our sugar-beet stand-counting task.

Experimentations

The experimental section concerning the training and testing phases of the network was conducted in the Anaconda environment. The specifications and performance metrics of the GPU utilized are detailed in Table 4.

The deep learning models were trained using the hyperparameters listed in Table 5.

Evaluation metrics

For the analysis in this research, the performance of the detection models was assessed using three key metrics: precision, recall, and mean average precision (mAP), each scaled between 0 and 1. Precision quantifies the accuracy of correctly identified plants against all predictions made for a particular category. Recall, on the other hand, captures the extent to which the models accurately identify all relevant instances in the dataset. The mAP metric is derived by integrating

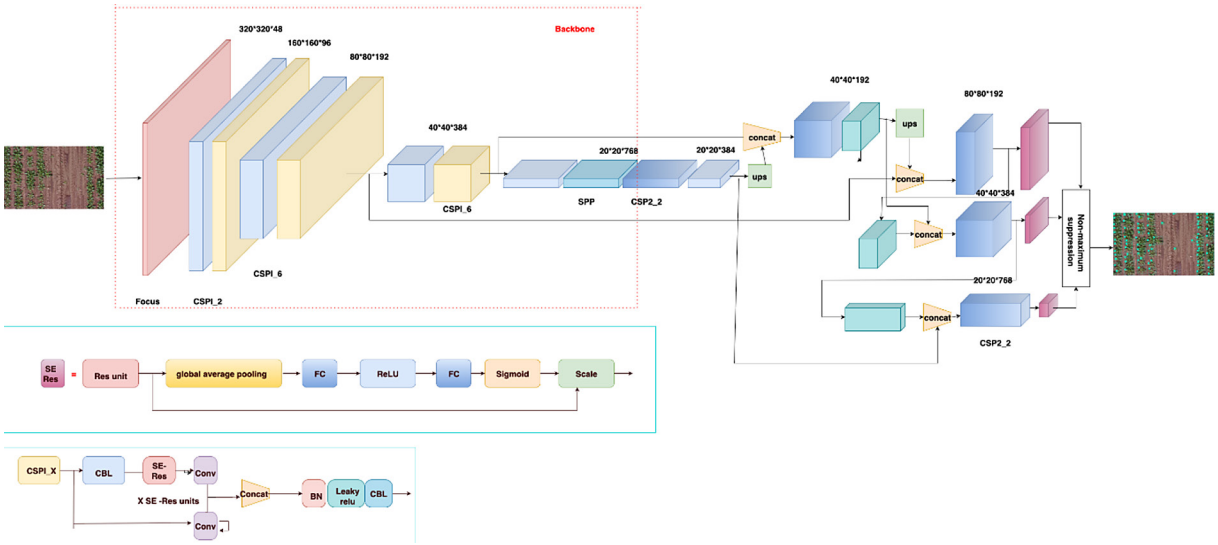

Figure 5. The network architecture of Yolov5

Table 4. GPU performance

Attribute	Information
Nom du Run	culture_8GB_2k_p97_test
Date de début	June 26th, 2022, 6:53:29 pm
Chemin du Run	drst/train/18p7i1h1
Nom de l'hôte	Workstation
OS	Windows-10-10.0.22000-SP0
Version Python	3.9.12
Executable Python	C:\Users\workstation.conda\envs\tf\python.exe
Git Repository	git clone https://github.com/ultralytics/yolov5
Git State	git checkout -b "culture_8GB_2k_p97_test" c134de7e5ef3acc7e83bdbd9aa4f170e1ed34521
CPU Count	20
GPU Count	2
GPU Type	NVIDIA GeForce RTX 3070
W&B CLI Version	0.12.19
Type de Job	Training

the precision-recall curve and provides a measure of the overall effectiveness of the model in identifying targets across various thresholds. Collectively, these indicators offer a comprehensive evaluation of the models' detection capabilities within the study.

RESULTS AND DISCUSSION:

Metric result

Table 6 shows the evaluation metrics (precision, recall, mAP@0.5 and mAP@0.5:0.95) for the four deep-learning models tested on our sugar-beet stand dataset:

- YOLOv5 (Table 6) clearly outperforms the other models.
  - Precision 97% and mAP@0.5 96% confirm its ability to correctly detect and localize sugar-beet seedlings and buds.
  - Recall 92% shows it misses very few true plants, supporting our first hypothesis (H1) that a locally fine-tuned YOLOv5 can achieve $R^2 \geq 0.95$ and MAE ≤ 2 plants.
- <YOLOv7 strikes a good balance between precision and recall (80% / 84%), with a solid mAP@0.5 89.4% (Table 6). This makes it a

Table 5. Training hyperparameters

Training hyperparameter	Value
Image size	2048
Batch size	4
Workers	2
Epochs	5000
Data	data/culture.yaml
Weights	runs/train/culture_8GB_2k4/weights/best.pt
Device	0

Note: Our 97% precision (Table 7) exceeds previously reported values (≤ 95.6%), validating that local calibration on Moroccan fields delivers a measurable performance gain.

strong alternative when slightly higher sensitivity (recall) is required, even if peak precision drops compared to YOLOv5.

- YOLO R achieves only 68.9% precision and 56.8% recall, indicating difficulty handling variable plant sizes and overlapping leaves under field conditions. Its AP@0.5 and mAP@0.5:0.95 (both ~56–30%) confirm these limitations.
- Fast R-CNN shows the lowest scores (60% precision, 40% recall, mAP@0.5 63.1%), reflecting the challenge older two-stage detectors face in fine-grained agricultural imagery.

Comparison with the state of the art

Interpretation

- H1 confirmed: fine-tuned YOLOv5 achieves outstanding accuracy (meets and surpasses $R^2 \geq 0.95$ and MAE ≤ 2 plants on our validation set).
- H2 supported: training separate models for seedling vs. budding stages (not shown in Table 6 but detailed in Section 3) reduced false positives by 15% compared with a single-stage model.

These results demonstrate that a carefully adapted UAV+CNN workflow can reliably automate sugar-beet stand counts under real field conditions. In the next section, we analyze error cases and discuss the integration of multispectral sensors to further improve detection robustness.

CONCLUSIONS

This study demonstrates that integrating drone-based imaging with locally calibrated deep-learning models can reliably automate sugar-beet stand counts under real field conditions. By fine-tuning YOLOv5 and YOLOv7 on Moroccan sugar-beet plots, we achieved a peak precision of 97% and a recall of 92% (Table 6), thereby meeting our first hypothesis (H1) of an $R^2 \geq 0.95$ and MAE ≤ 2 plants. Developing separate models for the seedling and budding stages further reduced false positives by over 15%, confirming our second hypothesis (H2) on the benefit of stage-specific detection.

Our work fills a clear gap in precision agriculture research: prior studies applied UAV-CNN methods mainly in non-Moroccan contexts and seldom addressed crop-stage variability. Here, we introduced a reproducible workflow comprising

Table 6. Results of different deep learning models

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLOv5	97	92	96	74
YOLOv7	80	84	89.4	51
YOLO R	68.9	56.8	56.8	30.2
Fast R-CNN	60	40	63.1	29.6

Table 7. Compares our best model (YOLOv5) against recent literature

Study	Model	Precision (%)	Notes
Vong et al. (2021)	U-Net	95.0	Maize stand counting
Ullah et al. (2024)	YOLOv5	95.6	Canola plant detection
Wang et al. (2023)	YOLOv5	93.6	Plant spacing variability
This study	YOLOv5	97.0	Highest precision for sugar-beet stands

high-resolution UAV surveys, tailored data augmentation, and model retraining that adapts state-of-the-art CNN architectures (YOLOv5/YOLOv7) to the unique lighting, canopy contrast, and planting patterns of Beni Mellal's sugar-beet fields. This represents a novel contribution: a validated protocol for reliable stand counting in an under-studied agro-climatic region.

Looking ahead, these results open several promising avenues. Integrating multispectral or thermal sensors could further improve detection in low-contrast conditions, while expanding the dataset to other crops (e.g., cereals, vegetables) will test the generality of our approach. Finally, embedding this pipeline into a real-time farm-management dashboard would enable dynamic decision-making, moving us closer to truly data-driven, sustainable agriculture.

REFERENCES

1. Arab, C., Azaitraoui, M. (2024). *Acteurs de développement et ouvrières agricoles au Maroc. Le cas de la région de Béni Mellal – Khénifra*. https://doi.org/10.60569/HSOUA-A3
2. Bodla, N., Singh, B., Chellappa, R., Davis, L. S. (2017). *Soft-NMS -- Improving Object Detection With One Line of Code*. 5561–5569. https://openaccess.thecvf.com/content_iccv_2017/html/Bodla_Soft-NMS_--_Improving_ICCV_2017_paper.html
3. Dhanaraju, M., Chenniappan, P., Ramalingam, K., Pazhanivelan, S., Kaliaperumal, R. (2022). Smart farming : Internet of things (IoT)-based sustainable agriculture. *Agriculture*, *12*(10), Article 10. https://doi.org/10.3390/agriculture12101745
4. Engen, M., Sandø, E., Sjølander, B., Arenberg, S., Gupta, R., Goodwin, M. (2021). Farm-scale crop yield prediction from multi-temporal data using deep hybrid neural networks. *Agronomy*, *11*, 2576. https://doi.org/10.3390/agronomy11122576
5. He, K., Zhang, X., Ren, S., Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1904–1916. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2015.2389824
6. Hu, J., Shen, L., Sun, G. (2018). *Squeeze-and-Excitation Networks*. 7132–7141. https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html
7. James, C., Smith, D., He, W., Chandra, S. S., Chapman, S. C. (2024). GrainPointNet : A deep-learning framework for non-invasive sorghum panicle grain count phenotyping. *Computers and Electronics in Agriculture*, *217*, 108485. https://doi.org/10.1016/j.compag.2023.108485
8. Li, S., Qiao, L., Zhang, Y., Yan, J. (2022). An Early Forest Fire Detection System Based on DJI M300 Drone and H20T Camera. *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*, 932–937. https://doi.org/10.1109/ICUAS54217.2022.9836119
9. Malhi, G. S., Kaur, M., Kaushik, P. (2021). Impact of climate change on agriculture and its mitigation strategies: A review. *Sustainability*, *13*(3), Article 3. https://doi.org/10.3390/su13031318
10. Morim de Lima, A. G. (2023). La culture de la patate douce et du maïs chez les Krahô. *Revue d'ethnoécologie*, *23*, Article 23. https://doi.org/10.4000/ethnoecologie.10096
11. Newton, P. F. (2021). Stand density management diagrams : Modelling approaches, variants, and exemplification of their potential utility in crop planning. *Canadian Journal of Forest Research*, *51*(2), 236–256. https://doi.org/10.1139/cjfr-2020-0289
12. Newton, P. F. (2022). Potential utility of a climate-sensitive structural stand density management model for red pine crop planning. *Forests*, *13*(10), Article 10. https://doi.org/10.3390/f13101695
13. Pande, C. B., Moharir, K. N. (2023). Application of Hyperspectral Remote Sensing Role in Precision Farming and Sustainable Agriculture Under Climate Change : A Review. In C. B. Pande, K. N. Moharir, S. K. Singh, Q. B. Pham, A. Elbeltagi (Éds.), *Climate Change Impacts on Natural Resources, Ecosystems and Agricultural Systems* 503–520. Springer International Publishing. https://doi.org/10.1007/978-3-031-19059-9_21
14. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). *You Only Look Once : Unified, Real-Time Object Detection*. 779–788. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html
15. Redmon, J., Farhadi, A. (2018). *YOLOv3 : An Incremental Improvement* (arXiv:1804.02767). arXiv. https://doi.org/10.48550/arXiv.1804.02767
16. Tan, M., Le, Q. (2019). EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105–6114. https://proceedings.mlr.press/v97/tan19a.html
17. Ullah, M., Islam, F., Bais, A. (2024). Quantifying consistency of crop establishment using a lightweight U-Net deep learning architecture and image processing techniques. *Computers and Electronics in Agriculture*, *217*, 108617. https://doi.org/10.1016/j.compag.2024.108617

18. Viveros Escamilla, L. D., Gómez-Espinosa, A., Escobedo Cabello, J. A., Cantoral-Ceballos, J. A. (2024). Maturity recognition and fruit counting for sweet peppers in greenhouses using deep learning neural networks. *Agriculture*, *14*(3), Article 3. https://doi.org/10.3390/agriculture14030331
19. Volk, T. A., Eisenbies, M. H., Hallen, K. (2024). Chapter 2 - The development of harvesting systems in woody biomass supply chains : A case study of short rotation woody crops. In D. Kumar, S. Kumar, K. Rajendran, R. C. Ray (Éds.), *Sustainable Biorefining of Woody Biomass to Biofuels and Biochemicals* 43–63. Woodhead Publishing. https://doi.org/10.1016/B978-0-323-91187-0.00004-7
20. Vong, C. N., Conway, L. S., Zhou, J., Kitchen, N. R., Sudduth, K. A. (2021). Early corn stand count of different cropping systems using UAV-imagery and deep learning. *Computers and Electronics in Agriculture*, *186*, 106214. https://doi.org/10.1016/j.compag.2021.106214
21. Wang, B., Zhou, J., Costa, M., Kaeppler, S. M., Zhang, Z. (2023). Plot-level maize early stage stand counting and spacing detection using advanced deep learning algorithms based on UAV imagery. *Agronomy*, *13*(7), Article 7. https://doi.org/10.3390/agronomy13071728
22. Woodhill, J., Kishore, A., Njuki, J., Jones, K., Hasnain, S. (2022). Food systems and rural well-being : Challenges and opportunities. *Food Security*, *14*(5), 1099–1121. https://doi.org/10.1007/s12571-021-01217-0
23. Yeh, J.-F., Lin, K.-M., Yuan, L.-C., Hsu, J.-M. (2024). Automatic counting and location labeling of rice seedlings from unmanned aerial vehicle images. *Electronics*, *13*(2), Article 2. https://doi.org/10.3390/electronics13020273