The Industry of the Future: From Industry 4.0 to Industry 5.0 – Integration of Humans and Technology: New Technologies

Karabegović, Isak

2025

Akademija nauka i umjetnosti Bosne i Hercegovine

https://bastina.anubih.ba/handle/123456789/837

DOI: 10.5644/PI2025.220.14

Accelerating Innovation in Healthcare through High-Performance Computing: Applications, Challenges and Future Perspectives

Madžida Hundur Hiyari*[1], Nejra Merdović[1], Merima Smajlhodžić Deljo[1],
Lemana Spahić[1], Basil Bošnjak[1], Lejla Gurbeta Pokvić[1,2]

Abstract: *Accurate estimation of wheat yield is essential for ensuring food security, especially given wheat's role in providing around 20 % of global calories and protein. Traditional yield estimation often relies on manual counting of wheat ears, a method that is labour-intensive, time-consuming, and impractical for large-scale production. To address these limitations, modern agriculture is increasingly turning to advanced technologies such as remote sensing, drones, and machine learning, which enable more efficient and precise monitoring of crop growth and yield potential.In this context, the present study introduces an automated ear-counting approach that applies machine learning to high-resolution images captured by unmanned aerial vehicles (UAVs). Drone imagery was collected during the late growth stage from 15 wheat fields in Bosnia and Herzegovina and processed at a resolution of 1024 × 1024 pixels. Images were manually annotated to mark regions containing wheat ears, resulting in a curated dataset of 556 high-resolution images. For detection, state-of-the-art models including Faster R-CNN, YOLOv8, and RT-DETR were used. While lower-quality images slightly reduced detection accuracy, overall model performance remained strong. This research demonstrates the value of combining UAV-based imaging with machine learning to modernise agricultural practices, offering an efficient, scalable solution for yield prediction and supporting greater sustainability and competitiveness in wheat production.*

Keywords: *precision agriculture; wheat ear counting; artificial intelligence; computer vision; real-time processing; detection*

1. Introduction

Wheat provides about 20 % of global calories and protein, making it essential for food security, particularly in developing regions. The Food and Agriculture Organization of the United Nations's 2030 agenda prioritises ending hunger and achieving food security, which requires investments in advanced agricultural technologies [1]. Accurate wheat yield forecasting supports better farm management and resource use but is complex due to factors like climate, soil,

---

*[1]Verlab Research Institute, Sarajevo, Bosnia and Herzegovina
[2] International Burch university, Sarajevo, Bosnia and Herzegovina
E-mail: ad@verlabinstitute.com

water availability, and farming practices. Traditional methods for counting wheat ears (e.g. using 1 m² frames) are labour-intensive and impractical on large fields. Automating this process can improve efficiency and accuracy. Recent advances include drones, sensors, and machine learning. Unmanned aerial vehicles (UAVs) equipped with multispectral cameras and sensors enable real-time crop monitoring, classification, and yield estimation, reducing costs and environmental impact [2-5] They also support tasks such as precision spraying, weed control, and water stress monitoring [6-11].

Bosnia and Herzegovina has underused agricultural land and suitable climate, but wheat yields are below national and European averages. Challenges include outdated techniques and limited technology adoption. Modernising production requires integrating advanced technologies. The shift to Agriculture 4.0 focuses on sustainability and automating tasks to boost efficiency [12-14]. Research has demonstrated machine learning algorithms with high precision in analysing crop images, such as TWSVM and deep learning models like PyTorch-based frameworks [15-16].

This study develops an automated wheat ear detection system using UAV imagery and computer vision models (Faster R-CNN, YOLOv8, RT-DETR) to improve accuracy over manual counting. The work aims to modernise farming practices in Bosnia and Herzegovina, supporting sustainable development and food security.

2. Materials and Methods

2.1. Locations

This study covers three locations in Bosnia and Herzegovina – Derventa, Odžak, and Goražde – where data were collected from 15 carefully selected fields of varying sizes (0.3–45.5 ha) and different wheat varieties. The locations were chosen to capture the geographic and climatic diversity of the country, from the lowland areas in the north to the hilly southeast. Although not all microclimatic and soil types are represented, the selected fields reflect key agricultural zones, making the results relevant for yield prediction and agricultural planning.

2.2. Data Collection

For the purposes of this study, data were collected using a digital camera and a drone. A PVC frame measuring 1 × 1 m was used with the drone, while a 0.5 × 0.5 m frame was used with the camera. This resulted in two separate datasets—one consisting of drone images and the other of images captured with the camera.

Figure 1. Example image taken by drone at height of 5m

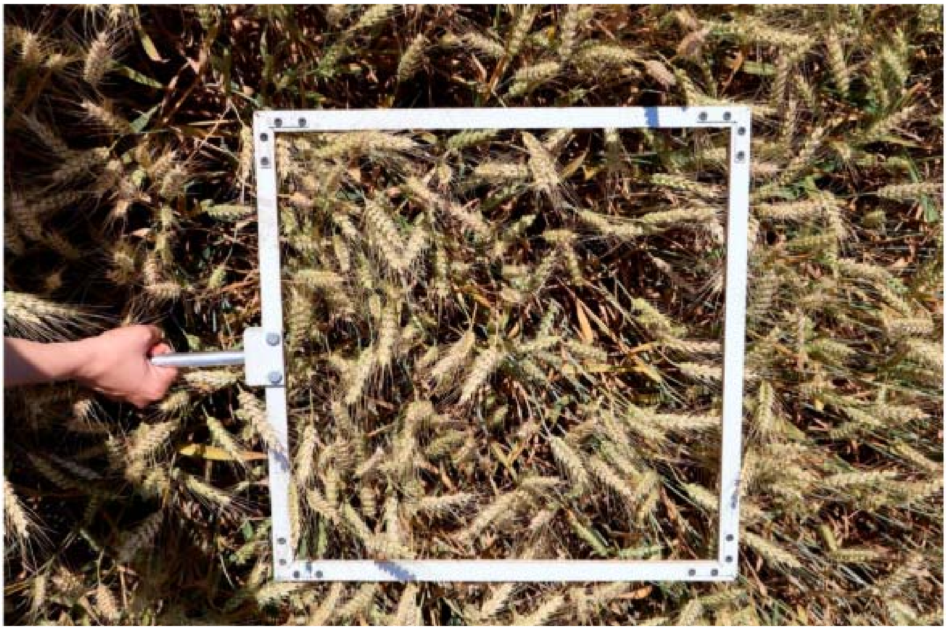

Figure 2. Example image taken by handheld digital camera

A DJI Mavic 3 Pro Cine drone equipped with a 20 MP RGB camera and HDR capability was used for aerial imaging, while a Canon EOS R10 camera was used for handheld imaging. For each location, 4 to 6 sample points were selected to reflect variations in wheat density, weed presence, and other factors. The drone captured images from heights of 5 m, 10 m, and 20 m (Figure 1.), while the handheld camera introduced variations in height and angle due to manual operation (Figure 2.).

2.1. Data Preprocessing

After collection, the images were transferred from the drone's internal memory (which is volatile and prone to data loss) to more secure locations—an external SSD as a physical backup and a cloud repository as the main storage for long-term use. This ensured the long-term availability and safety of the data required for training the AI model.

The original drone images were captured in high resolution (5280 × 2970 pixels), 96 dpi, and 24-bit color depth, enabling a high level of detail necessary for accurate object recognition. However, further processing was required to prepare them for use in model training. The first step in the processing was the extraction of the PVC frame from each image. Since the frame position varied from image to image and was not fixed, this step could not be automated and was instead performed manually.

During this process, the parts of the images containing the PVC frame—which defined the sample area in the field—were selectively cropped. Figure 3.a) shows the original image with a clearly marked blue frame (indicating the region of interest), while Figure 3.b) shows the cropped image used for further processing.

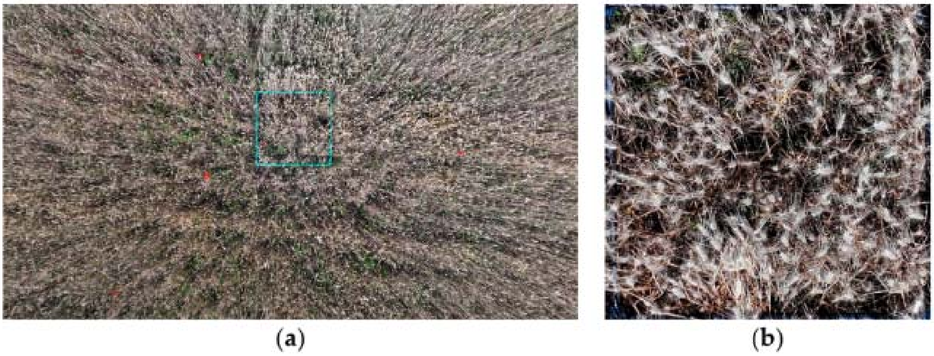

Figure 3.a) Original image take by arial drone; b) scaled image with extracted frame

After frame extraction, the images were resized to dimensions of 2048 × 2048 pixels to standardize the input size. These images were then divided into four quadrants, each 1024 × 1024 pixels in size. This division facilitated the annotation process, as smaller image segments allowed for more precise detection and labeling of wheat ears.

The complete image processing workflow—from image capture to annotation—is presented in the diagram below.

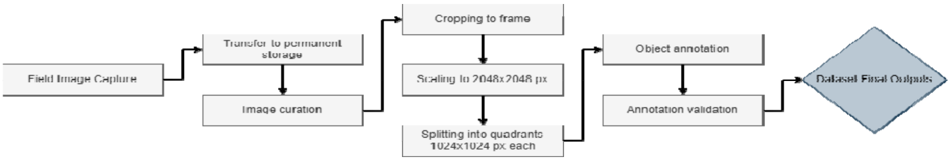

Figure 4. (a) Flowchart diagram depicting steps for image processing

Image annotation involved marking regions of interest (ROI) that contained the object to be detected—in this case, a wheat ear. Typically, each ROI contained only one wheat ear, and each was labeled with the tag “wheat ear.” No other objects (e.g., weeds, flowers, soil) were annotated in the dataset, to keep the model focused solely on wheat ear detection.

The annotation process was carried out entirely manually by human annotators. To ensure the accuracy and consistency of the dataset, a two-person review protocol was implemented. One person performed the initial annotation, while another independently reviewed each image, checked the accuracy of the labeled wheat ears, and made corrections as needed. This two-step quality control system significantly reduced the possibility of errors and improved the reliability of the dataset used for training the AI model.

2.1. Models for Object Detection

In this study, we utilized three state-of-the-art object detection models—Faster R-CNN [17], YOLOv8 [18] and RT-DETR [19] to detect wheat ears in RGB drone images. These models were chosen for their complementary strengths in terms of accuracy, speed, and ability to handle varying conditions.

Faster R-CNN is a two-stage detector that combines region proposal generation and object classification. It uses a Region Proposal Network (RPN) to generate candidate object regions and a CNN backbone (in our case, ResNet-50 with FPN) for feature extraction. Though highly accurate, especially in complex scenes, it is computationally demanding and not suitable for real-time applications.

YOLOv8, part of the "You Only Look Once" family, is a single-stage detector optimized for real-time performance. It uses a CSP-based backbone and combines FPN and PANet structures to extract multi-scale features. YOLOv8 is efficient and versatile, supporting object detection, classification, and segmentation. We used the lightweight "nano" version (~3 million parameters) suitable for fast inference, although its performance may decrease in cluttered or low-contrast scenes.

RT-DETR (Real-Time Detection Transformer) is a transformer-based detector capable of modeling global context and detecting small objects effectively. It employs self-attention mechanisms and learnable object queries to predict bounding boxes and classes. While powerful, its reliance on large datasets and a higher parameter count (~43 million) may limit its use in resource-constrained settings.

All models were trained for 100 epochs on a stratified dataset split (56% training, 14% validation, 30% test). During initial tests, we also examined model performance on a smaller high-quality subset to evaluate the impact of image clarity on detection results. This multi-model approach allowed us to assess detection performance under different conditions and requirements.

3. Results

The main outcome of this study is the development of a comprehensive dataset for wheat ear detection that can be used to train artificial intelligence models for agricultural applications. A total of 556 high-resolution images were collected under varied field conditions using drones and handheld cameras. Each image was carefully annotated and organized by location (lot), capture height, and sensor type, ensuring a transparent structure and reproducibility in data processing. This dataset provides a foundation for building effective systems for crop monitoring and yield prediction.

After preprocessing, deep learning models (Faster R-CNN, RT-DETR, and YOLOv8) were trained for 100 epochs, with their accuracy assessed using standard metrics: precision, recall, and F1-score. Initial experiments were conducted on a smaller set of high-quality images (4 for training and 5 each for validation and testing) to evaluate the effect of image quality on model performance. Including lower-quality images led to a slight drop in metrics, particularly in recall, but did not significantly affect overall test results.

Figure 5. Image fragment with labels

Figure 6. Detection examples for low-quality images: (a) Faster R-CNN (F1:0.87), (b) RT-DETR (F1:0.77), (c) YOLOv8 (F1: 0.77)

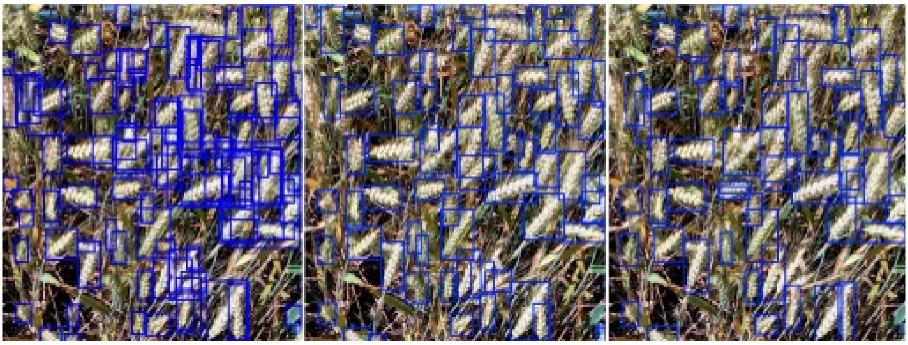

Figure 7. Detection examples for high-quality images: (a) Faster R-CNN (F1: 0.81), (b) RT-DETR (F1:0.74), (c) YOLOv8 (F1: 0.69)

For example, in early tests, RT-DETR achieved 0.77 precision and 0.72 recall, showing the best balance between accuracy and detection capability. By contrast, YOLOv8 tended to miss more true positive instances, while Faster R-CNN often produced multiple overlapping boxes in dense image regions, highlighting challenges in localizing individual wheat ears in complex conditions.

Full-dataset testing confirmed these trends. YOLOv8 achieved the highest precision (0.87), indicating few false positives, but had significantly lower recall (0.46), suggesting it often failed to detect all wheat ears present in an image. RT-DETR achieved the highest F1-score (0.69), thanks to its attention mechanism and two-stage detection process, which enable better detection of objects of varying sizes, including small and partially occluded wheat ears.

While YOLOv8 is well-suited for applications requiring high speed and precision, RT-DETR proved to be the most reliable model for wheat ear detection across diverse field data conditions, making it suitable for integration into smart agriculture systems.

Table 1. Evaluation metrics for initial experiments with high-quality training images

Num.of High/Low-Quality Images	**Model**	**Precision**	**Recall**	**F1 Score**
4/0	Faster R-CNN	0.69251	0.69299	0.6932
	RT-DETR	0.77081	0.71687	0.743
	YOLOv8	0.66957	0.67749	0.6736
4/2	Faster R-CNN	0.67938	0.6988	0.6885
	RT-DETR	0.74461	0.7259	0.7336
	YOLOv8	0.66901	0.6575	0.6613
4/4	Faster R-CNN	0.6844	0.6534	0.6691
	RT-DETR	0.762	0.63392	0.6926
	YOLOv8	0.64574	0.6015	0.6232

Table 2. Evaluation metrics for final models

	Precision	**Recall**	**F1 Score**
Faster R-CNN	0.60209	0.57907	0.59036
RT-DETR	0.68422	**0.70811**	**0.69596**
YOLOv8	**0.87257**	0.46132	0.60355

4. Discussion

Wheat yield prediction systems leverage various data sources and machine learning techniques to forecast crop yields accurately and in a timely manner. This synthesisAgriEngineering 2024, 6 4717 examines the effectiveness of different wheat yield prediction systems based on recent research findings. Through the development of automated systems for tasks like wheat ear detection, this study has shown how UAVs, high-resolution cameras, and deep learning models can drastically improve efficiency, accuracy, and scalability compared to traditional manual methods. By utilising large datasets collected from diverse geographical locations and processing them using state-of-the-art AI models, such as Faster R-CNN, YOLOv8, and RT-DETR, significant strides were made in object detection and precision agriculture. It can be observed from Table 2 that the YOLOv8 model achieves the highest precision, indicating that this model is the most effective for minimising false positives. However, its recall is considerably lower, suggesting the model misses more than a half of labelled wheat ears. The RT-DETR model has the highest recall, meaning it is more effective at identifying true positives but with slightly lower precision than YOLOv8. The results show that there are strengths and weaknesses among the models used in the study. Another factor that is worth considering is that the dataset used included images from fields of various sizes, altitudes, and wheat species sown. This enables these models to be able to detect wheat ears in varied conditions. However, this approach emphasises the importance of a high-quality dataset as well as the size of the dataset. The results achieved in this study demonstrate an improvement compared to those from two previously discussed studies [15,20], where the highest precision for wheat ear detection reached 82%. Unlike these prior works, which focused on datasets with limited variability in terms of wheat species, field conditions, and image quality, this study utilised a more diverse dataset that included images captured under varied conditions such as different field sizes, altitudes, and wheat species. This diversity enhanced the generalisability of the models, enabling them to perform well across a broader range of scenarios. Future improvements should address the limitations observed in the current results, particularly the low recall of YOLOv8, which indicates missed detections due to dense or overgrown wheat ears. Given that the collected data span various climatic conditions and include wheat of different varieties and soil types, our future work will focus on examining how these factors influence the models. In this study, we concentrated on the general task of detecting wheat ears; however, we are particularly interested in investigating whether the models exhibit any bias towards specific wheat varieties or climatic conditions in which the wheat grows. This research highlights that high-quality datasets combined with robust deep learning algorithms enable reliable models for crop monitoring. The integration of artificial intelligence automates labour-intensive tasks, facilitates timely decision-making, and improves yield forecasting.

5. Conclusion

The application of digital technologies, such as computer vision and artificial intelligence, represents a powerful tool in precision agriculture, particularly for crop detection and yield estimation. The performance of algorithms, with a precision of 0.87257 achieved with YOLOv8, a recall of 0.70811 with RT-DETR, and an F1 score of 0.69596, highlight theAgriEngineering 2024, 6 4718 potential deep learning has when it comes to wheat ear detection. This study lays the groundwork for future research, not only for wheat production but also for other crops and agricultural products. These technologies can be used for precise fruit counting, plant species identification, crop growth and development monitoring, and pest and disease detection at early stages, reducing the need for chemical treatments and enhancing resource efficiency. Additionally, they have potential applications in assessing the quality of fruits and vegetables, optimising irrigation through soil moisture analysis, and estimating yields for various types of grains, fruits, vegetables, and industrial crops. In the future, such systems could be integrated into broader farm management platforms, providing farmers with personalised recommendations based on data collected from drones, satellites, and other IoT devices. This would enable not only resource optimisation and yield improvement but also represent a significant contribution to global food security through sustainable and competitive agricultural practices.

4. References

[1] Food and Agriculture Organization of the United Nations. Available online: https://www.fao.org/sustainable-development-goals-helpdesk/en (accessed on 23 July 2024).

[2] Saeed, K.; Lizhi, W. Crop Yield Prediction Using Deep Neural Networks. Front. Plant Sci. 2019, 10, 621.

[3] Jones, E.J.; Bishop, T.F.; Malone, B.P.; Hulme, P.J.; Whelan, B.M.; Filippi, P. Identifying causes of crop yield variability with interpretive machine learning. Comput. Electron. Agric. 2022, 192, 106632. [CrossRef]

[5] Ma, J.; Wu, Y.; Liu, B.; Zhang, W.; Wang, B.; Chen, Z.; Wang, G.; Guo, A. Wheat Yield Prediction Using Unmanned Aerial Vehicle RGB-Imagery-Based Convolutional Neural Network and Limited Training Samples. Remote Sens. 2023, 15, 5444. [CrossRef]

[6] Grbovi´c, Ž.; Pani´c, M.; Marko, O.; Brdar, S.; Crnojevic, V. Wheat Ear Detection in RGB and Thermal Images Using Deep Neural Networks. In Proceedings of the International Conference on Machine Learning and Data Mining, MLDM, New York, NY, USA, 13–18 July 2019.

[7] Istiak, M.A.; Syeed, M.M.; Hossain, M.S.; Uddin, M.F.; Hasan, M.; Khan, R.H.; Azad, N.S. Adoption of Unmanned Aerial Vehicle (UAV) imagery in agricultural management: A systematic literature review. Ecol. Inform. 2023, 78, 102305. [CrossRef]

[8] Enciso, J.; Avila, C.A.; Jung, J.; Elsayed-Farag, S.; Chang, A.; Yeom, J.; Landivar, J.; Maeda, M.; Chavez, J.C. Validation of Agro-nomic UAV and Field Measurements for Tomato Varieties. Comput. Electron. Agric. 2019, 158, 278–283. [CrossRef] AgriEngineering 2024, 6 4719

[9] Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images. Clust. Comput. 2023, 26, 1297–1317. [CrossRef] [PubMed]

[10] Surekha, P.; Venu, N.; Shetty, A.N.; Sachan, O. An Automatic Drone to Survey Orchards Using Image Processing and Solar Energy. In Proceedings of the IEEE 17th India Council International Conference (INDICON), New Delhi, India, 10–13 December 2020; pp. 1–7. [CrossRef]

[11] Brewer, K.; Clulow, A.; Sibanda, M.; Gokool, S.; Odindi, J.; Mutanga, O.; Naiken, V.; Chimonyo, V.G.P.; Mabhaudhi, T. Estimation of Maize Foliar Temperature and Stomatal Conductance as Indicators of Water Stress Based on Optical and Thermal Imagery Acquired Using an Unmanned Aerial Vehicle (UAV) Platform. Drones 2022, 6, 169. [CrossRef]

[12] Foreign Investment Promotion Agency of Bosnia and Herzegovina. Available online:

http://www.fipa.gov.ba/atraktivni_sektori/poljoprivreda/default.aspx?langTag=en-US (accessed on 2 September 2024).

[13] The Parliament of the Federation of Bosnia and Herzegovina. Strategija Razvoja FBiH 2021-2027_bos. Available online: https://parlamentfbih.gov.ba (accessed on 23 July 2024).

[14] World Bank. Available online:

[15] Zhou, C.; Liang, D.; Yang, X.; Yang, H.; Yue, J.; Yang, G. Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. Front. Plant Sci. 2018, 9, 1024. [CrossRef] [PubMed]

[16] Haq SI, U.; Tahir, M.N.; Lan, Y. Weed detection in wheat crops using image analysis and artificial intelligence (AI). Appl. Sci. 2023,13, 8840. [CrossRef]

[17] Jabir, Brahim & El Moutaouakil, Khalid & Noureddine, Falih. (2023). Developing an Efficient System with Mask R-CNN for Agricultural Applications. Agris on-line Papers in Economics and Informatics. 15. 61-72. 10.7160/aol.2023.150105.

[18] Kanna, Kamalesh & Ramalingam, Kumaraperumal & P, Pazhanivelan & Ramasamy, Jagadeeswaran & Pc, Prabu. (2024). YOLO deep learning algorithm for object detection in agriculture: a review. Journal of Agricultural Engineering. 55. 10.4081/jae.2024.1641.

[19] García-Navarrete, O.L.; Camacho-Tamayo, J.H.; Bregon, A.B.; Martín-García, J.; Navas-Gracia, L.M. Performance Analysis of Real-Time Detection Transformer and You Only Look Once Models for Weed Detection in Maize Cultivation. Agronomy 2025, 15, 796. https://doi.org/10.3390/agronomy15040796

[20] Haq SI, U.; Tahir, M.N.; Lan, Y. Weed detection in wheat crops using image analysis and artificial intelligence (AI). Appl. Sci. 2023, 13, 8840. [CrossRef]