A real-time intelligent framework for wheat stripe rust management using lightweight deep learning and LLMs

Yi-Teng Wu [a], Qi-Yuan Zhang [a], Hao-Jie Chen [a], Zhan-Hong Ma [b], Wen-Hao Su [a,*]

[a] College of Engineering, China Agricultural University, 17 Qinghua East Road, Haidian, Beijing 100083, China

[b] College of Plant Protection, China Agricultural University, 2 Yuanmingyuan West Road, Haidian, Beijing 100193, China

* Correspondence: wenhao.su@cau.edu.cn

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract.

Wheat stripe rust, caused by Puccinia striiformis f. sp. tritici, remains a major constraint on wheat production, and reliable severity grading is essential for timely, site-specific disease management. However, practical field monitoring still faces two linked challenges: visual models must maintain sufficient accuracy under complex backgrounds while running on low-power edge devices, and diagnostic outputs must be translated into actionable management recommendations rather than remaining as isolated classification results. To address these limitations, this study developed an integrated edge-intelligence framework that combines lightweight visual perception with large language model (LLM)-assisted cognitive reasoning for wheat stripe rust diagnosis and management. The system first applies YOLO-SSC, an improved YOLOv11n-seg instance segmentation model incorporating Space-to-Depth Convolution, Swin Transformer features, and Context Aggregation, to isolate wheat leaf regions and suppress background interference. The resulting cropped leaf images and masks are then processed by WheatSeverityNet, a compact mask-guided convolutional network designed for six-level disease severity grading through multi-scale feature extraction, lightweight attention, and mask-aware global pooling. Finally, the structured visual outputs are combined with crop growth stage, meteorological conditions, and pesticide-use history, and passed to a DeepSeek LLM module to generate diagnostic interpretation, urgency assessment, control recommendations, and prognosis-oriented guidance. Experimental results showed that YOLO-SSC achieved 98.85% mAP_{50} and 80.27 FPS on an RTX 4080 workstation, while reducing parameters by 16.35% and GFLOPs by 11.76% compared with

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

the baseline. WheatSeverityNet achieved 83.38% test accuracy with only 0.124 M parameters and a model size of 0.54 MB. Both models were deployed on a Raspberry Pi 5 handheld terminal, where the system achieved 4.59 FPS during real-time video preview and less than 200 ms image-to-result latency in CAPTURE mode. These results demonstrate that the proposed framework can support portable, real-time wheat stripe rust assessment while extending disease monitoring from visual perception to context-aware assisted decision-making. Future work will focus on larger multi-regional validation, improved robustness under dense canopy occlusion, and field evaluation of the agronomic effectiveness of LLM-guided recommendations.

Keywords.
Disease severity grading; Instance segmentation; Lightweight network; Large language model; Edge computing

Introduction

Wheat stripe rust (Puccinia striiformis f. sp. tritici) can cause serious yield losses, and severity grading rather than simple disease detection is the information most directly needed for precision control (Jiang et al., 2023; Sarkar et al., 2023). Manual field surveys are slow and subjective, whereas many deep-learning systems either depend on desktop-class computing or stop at a visual label that still requires expert interpretation. The objective of this study was therefore to develop a portable system that links field image acquisition, leaf segmentation, severity grading, and context-aware management reasoning in one workflow.

Materials and Methods

The dataset combined 990 self-collected wheat stripe rust images from the Shangzhuang Experimental Station of China Agricultural University with 894 manually screened web images. Images were divided into training, validation, and test sets at a 7:2:1 ratio, and only the training and validation sets were augmented. Fine-grained annotations were produced with X-AnyLabeling and SAM 2.1 assistance. A second grading dataset of 2,361 images was built from high-quality samples and the Yellow Rust 19 public dataset, with six severity levels defined by the lesion-area ratio.

The proposed visual pipeline contains two serial models (Fig. 1). YOLO-SSC uses SPDConv to reduce redundant downsampling, Swin Transformer blocks to enhance global context, and Context Aggregation to strengthen multi-scale representation. WheatSeverityNet receives the cropped RGB leaf and its mask, then applies lightweight multi-scale feature extraction, channel-spatial attention, and mask-aware global pooling so that background pixels contribute minimally to severity discrimination. Both networks were exported to ONNX and deployed with ONNX Runtime on an 8 GB Raspberry Pi 5 handheld terminal equipped with an 8-megapixel CSI camera.

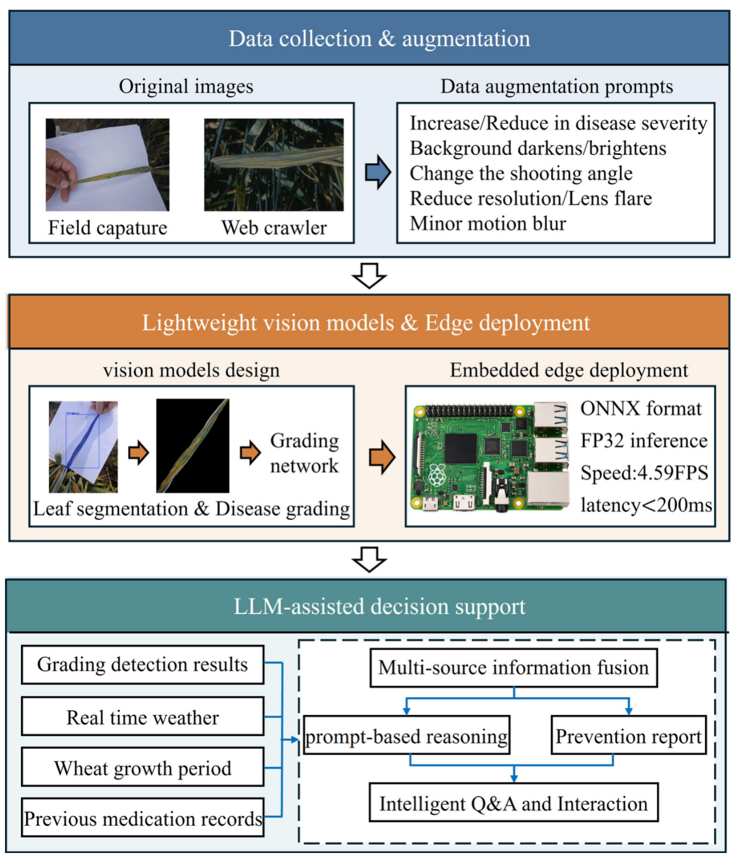

Fig. 1. System pipeline for wheat stripe rust management, including data acquisition, lightweight visual perception, Raspberry Pi 5 edge deployment, and LLM-assisted decision support.

For management support, the structured outputs of the vision models, together with user-confirmed severity, crop growth stage, meteorological data, and pesticide-use history, were passed to a DeepSeek LLM module. Prompt constraints guided the model to return diagnostic analysis, urgency assessment, medication recommendations, control measures, and prognosis prediction in a consistent format while supporting multi-turn consultation.

Results and Discussion

YOLO-SSC achieved an effective accuracy-efficiency balance. Compared with the YOLOv11n-seg baseline, it reduced parameters from 2.83 M to 2.37 M and lowered GFLOPs by 11.76%, while mAP_{50} increased by 1.68% and 0.98% on two independent test sets. The final workstation performance reached 98.85% mAP_{50} and 80.27 FPS, indicating that the added global and contextual modules improved robustness without sacrificing real-time capability.

WheatSeverityNet obtained 83.38% test accuracy with only 0.124 M parameters and a 0.54 MB model size. Its errors were mainly between adjacent severity levels, which is consistent with the continuous nature of disease progression. Mask-quality sensitivity analysis showed that grading accuracy remained above 80% when upstream mask IoU was at least 0.70, confirming the importance of reliable leaf segmentation. On the Raspberry Pi 5, the serial system reached 4.59 FPS in video preview and less than 200 ms latency in CAPTURE mode, which is adequate for handheld scouting. The LLM module further extends the system from perception to assisted decision-making by translating disease grade and field context into management-oriented recommendations.

Conclusion

This short paper presents an edge-intelligence framework for wheat stripe rust management. By combining YOLO-SSC, WheatSeverityNet, Raspberry Pi 5 deployment, and LLM-assisted reasoning, the system supports rapid field diagnosis and context-aware recommendations. Future work should expand multi-regional validation, improve robustness under dense canopy occlusion, and evaluate the long-term agronomic value of LLM-guided recommendations in field trials.

Acknowledgments

The authors gratefully acknowledge the support of China Agricultural University for providing experimental conditions and field resources. We also thank the members of the research team who assisted with image collection, annotation, model testing, and field device evaluation.

References

Hayit, T., Erbay, H., Varcin, F., Hayit, F., & Akci, N. (2021). Determination of the severity level of yellow rust disease in wheat by using convolutional neural networks. Journal of Plant Pathology, 103(3), 923-934.

Jiang, Q., Wang, H., & Wang, H. (2023). Severity assessment of wheat stripe rust based on machine learning. Frontiers in Plant Science, 14, 1150855.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision.

Sarkar, C., Gupta, D., Gupta, U., & Hazarika, B. B. (2023). Leaf disease detection using machine learning and deep learning: Review and challenges. Applied Soft Computing, 145, 110534.

Sunkara, R., & Luo, T. (2022). No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. ECML PKDD.