International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 11 | Nov 2026 www.irjet.net p-ISSN: 2395-0072

Agentic AI and Multi-Agent RAG for Intelligent Decision Support: A Survey and Conceptual Framework for Grain Quality Assessment

Abhang Mali[1], Arti Bhise[2], Madan Mali[3]

[12]Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune
[3]School of Engineering and Technology, DES Pune University, Pune

---

Abstract - *With the advent of Artificial Intelligence (AI), the hard work required in the day-to-day tasks of humans has been reduced to a greater extent. Different technical and non-technical tasks are now largely being transformed into intelligent AI systems. Due to the deep analysis of many researchers and developers, Agentic AI systems have now gained high adoption in many fields. Similarly, the Retrieval Augmented Generation (RAG) framework has gained significant attention for enhanced reliability of the AI-generated responses. This paper studies the concepts, architecture, working principles, applications, advantages, and limitations of Agentic AI and Multi-Agent RAG. It mainly focuses on the joint working of multi-agents of Agentic AI with the RAG framework. The paper represents a comprehensive study of researchers who have worked on this. The study aims to provide a clear understanding of Multi-Agent RAG systems and their capability in decision support systems.*

Keywords: Artificial Intelligence (AI), Agentic AI, Intelligent Decision Support, Grain Quality, Multi-agent, Retrieval-Augmented Generation (RAG).

1. INTRODUCTION

Artificial intelligence (AI) systems are used to automate human tasks. The transition from rules-based AI to generative AI has played a significant role in these years. Different efficient generative models like Generative Adversarial Networks (GANs) [1], diffusion models [2], transformers [3], etc, helped in explicit content generation. These AI systems thus evolved by generating not only text, but other data types like images, videos, audios, and speech. One of the drawbacks of these AI systems is the inability to represent factual correctness. The systems cannot fully ensure whether the generated content are reliable. This drawback matters in different important fields like agriculture, healthcare, finance, banking, organization management, and many more.

To address this limitation, the Retrieval-Augmented Generation (RAG) framework [4] has emerged as an effective approach. This framework retrieves relevant information from external knowledge sources. For example, for booking a flight ticket, the system used to learn the flight booking details and from the official website and then provide the legitimate response. Due to the use of external sources as a knowledge base, these RAG-based AI systems have increased their adoption in various real-life applications. One of the major fields is of product quality assessment. Conventional AI-based computer vision systems used to detect or segment and then provide the respective impurity identification result. With the help of the RAG framework, AI systems now detect and use external information about that product before finalizing their response. This increases reliability and customer satisfaction.

Furthermore, Agentic AI has evolved at a greater pace. It plans, uses external tools, reasons, and takes final decisions. Agentic AI's multi-agent framework that includes the communication and collaboration of different agents for one complex problem has significantly reduced the complexity of the operation. The integration of Agentic AI and RAG has become a milestone in the field of decision support systems. Therefore, studying the architectures, working principles, applications, advantages, and limitations of Agentic AI and Multi-Agent RAG is important.

2. LITERATURE REVIEW

Many researchers and developers have worked on the Agentic AI and RAG framework. They have proposed new algorithms or workflows to ease the automation task. The development of Retrieval-Augmented Generation (RAG) has provided an important approach to improving the reliability of responses generated by the LLMs. [4] introduced a RAG architecture that combines a pre-trained sequence-to-sequence generator with a dense external document retriever. The approach enables the LLM to retrieve relevant information from an external knowledge source and use it as additional context when generating responses. It was found that RAG-generated responses were considered more factual than Bidirectional and Auto-Regressive Transformer (BART) baseline responses in 42.7% of comparisons.

[5] proposed Reasoning and Acting (ReAct), a prompting framework that integrates reasoning traces with actions performed through external tools or environments. Unlike the conventional approaches that generate an answer directly, ReAct allows a language model to reason about a problem, execute an action, observe the resulting information, and use that observation to determine its next step. Their work highlighted how combining reasoning and external actions can support more structured task execution and improve the ability of language models to interact with external environments.

[6] presented a comprehensive survey of LLM-based autonomous agents. This included the agents' construction, capabilities, applications, evaluation strategies, and challenges. The study organized the development of autonomous agents around important components such as profiling, memory, planning, reasoning, and tool utilization. The research provided a clear understanding of the major components required to develop autonomous systems. Similarly, another survey was conducted by [7] on the LLM-Based Multi-Agent Systems: Workflow, Infrastructure, and Challenges. The study presented a workflow-oriented understanding of multi-agent systems. It included the study of how individual agents are structured, how they perceive information, perform actions, interact with other agents, and evolve during task execution. [8] further investigated the development of LLM-based multi-agent systems through their survey, Large Language Model Based Multi-Agents: A Survey of Progress and Challenges. The authors reviewed the evolution of multi-agent approaches. The focus was on agent capabilities, communication mechanisms, application domains, and evaluation methods. The study explained how multiple LLM-based agents can collaborate by distributing responsibilities and exchanging information to solve complex problems.

With the increasing interest in combining agent-based reasoning with external knowledge retrieval, [9] proposed the Multi-Agent Retrieval-Augmented Generation (MA-RAG) framework. This framework was designed to address complex information-seeking tasks. The proposed architecture consists of specialized Planner, Step Definer, Extractor, and Question Answering agents. These agents collaboratively perform query interpretation, task decomposition, information extraction, and answer generation. The framework uses intermediate reasoning and communication between agents to handle ambiguous queries and information distributed across multiple documents. The proposed framework outperformed training-free baseline approaches.

[10] introduced a framework called Multi-Agent Filtering Retrieval-Augmented Generation (MAIN-RAG). This framework focuses on improving the quality of retrieved information before generating a response. The framework employs multiple LLM-based agents to collaboratively evaluate, filter, and score retrieved documents. It also incorporates an adaptive filtering mechanism that adjusts the relevance threshold according to the distribution of document scores. This helps reduce irrelevant information while retaining useful evidence for response generation. As compared to the traditional RAG approaches, this proposed framework led to an increase in answer-generating accuracy from 2% to 11%. [11] presented a survey titled Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG, examining the integration of autonomous agents with retrieval-augmented generation. The study discussed different Agentic RAG architectures, design patterns, applications, and implementation strategies. It also examined how agents can dynamically plan retrieval operations, evaluate retrieved information, perform additional searches when required, and refine generated responses. Furthermore, the authors highlighted important challenges related to agent coordination, memory management, evaluation, and governance. This survey provides an overview of the transition from conventional RAG pipelines towards more adaptive and autonomous retrieval-based systems.

Further, [12] presented Ask in Any Modality: A Comprehensive Survey on Multimodal Retrieval-Augmented Generation. The study reviewed approaches that integrate different information modalities, including text, images, audio, and video, into retrieval-augmented generation systems. The authors discussed multimodal datasets, benchmarks, evaluation metrics, retrieval strategies, information fusion, and generation techniques. They also examined challenges associated with cross-modal alignment, multimodal reasoning, and robustness. This research is particularly relevant to applications that require combining visual information with external textual knowledge, such as image-based grain inspection supported by retrieved quality specifications. In the agricultural domain, [13] conducted a comprehensive survey titled Computer Vision Based Food Grain Classification: A Comprehensive Survey. The authors reviewed computer-vision-based approaches for identifying and classifying different food-grain varieties. Their analysis covered the major stages of grain classification pipelines, including image acquisition, preprocessing, feature extraction, and classification. The survey organized existing methods and highlighted important considerations for developing computer-vision-based grain identification and quality inspection systems.

[14] presented a review titled An Overview of the Current Situation and Future Development Direction of Grain Detection: Taking Computer Vision Combined with Deep Learning. It included an examination of the application of

computer vision and deep-learning techniques in grain detection and quality control. The study discussed the limitations of traditional visual inspection methods and described how automated image-processing techniques can support grain identification, defect detection, and quality evaluation. It also reviewed the use of deep-learning approaches in grain processing and discussed future development directions for automated inspection. [15] investigated the use of deep-learning algorithms and flatbed scanning technology for automated rice-quality assessment. The proposed approach identifies individual rice grains and detects impurities in scanned images for subsequent quality analysis. An important finding of the study was that the scanning setup enabled the identification of approximately 3,000 objects in a single scanner-ruler pass. This demonstrates the potential of combining image acquisition and deep-learning-based detection for grain inspection.

Further focusing on quantitative impurity assessment, the study by [16] included the development of a machine-vision approach for estimating impurity content in paddy grains. The proposed method incorporated YOLOv8n-DLMF-based instance segmentation to identify grain and impurity regions. The proposed method achieved a mean relative error of 6.02%, Since this work connects visual detection with quantitative impurity measurement, it is relevant for automated grain-quality evaluation. [17] proposed CottonBot, an AI-driven agricultural assistant that integrates LLMs, RAG, and Agentic AI tools to support cotton farming and irrigation decisions. The system includes domain-specific agricultural information, soil-moisture sensor readings, and weather forecasts to generate context-aware recommendations. Multiple embedding models and vector-store configurations were evaluated for document retrieval. The conducted experiments showed that the all-MiniLM-L6-v2 embedding model combined with FAISS achieved a Mean Reciprocal Rank (MRR) of 0.8855, Recall@5 of 0.9476, and Precision@5 of 0.9285. For response generation, Llama 3.1 (8B) achieved a faithfulness score of 0.9681 and cosine similarity of 0.9523, with an average generation time of 6.2 seconds. These findings explain how RAG and agentic tools can be integrated to support knowledge-grounded agricultural decision-making.

Another AI system was introduced by [18]. This was a domain-governed agricultural knowledge system. It integrated structured agricultural knowledge with RAG. The framework incorporates a Domain Knowledge Processing Layer (DKPL), which uses domain concepts, causal rules, and agronomic thresholds to guide document retrieval and validate generated responses. An agricultural knowledge corpus was constructed using 19 agricultural textbooks, and the system was evaluated using a benchmark containing 504 agricultural questions across four Food and Agriculture Organization (FAO) / Department of Agriculture (USDA) related knowledge categories. Their experiments showed that ChatGPT-4o Mini combined with DKPL-constrained RAG achieved an accuracy of 95.2%, compared with 93.6% using standard RAG. The proposed DKPL also reduced irrelevant or semantically misaligned retrieval by approximately 27–31%. These findings illustrate the benefits of leveraging structured domain knowledge and retrieval constraints into LLM-based agricultural decision-support systems.

Taken together, the reviewed studies demonstrate significant developments in retrieval-augmented generation, autonomous AI agents, multi-agent collaboration, multimodal information processing, and computer-vision-based agricultural inspection. Research on Multi-Agent RAG approaches demonstrates the collaboration of specialized agents and their respective improvement in retrieval and response generation. In parallel, computer-vision studies have demonstrated the feasibility of grain classification, impurity detection, and quantitative quality assessment. Agricultural applications such as CottonBot and AgroLLM further illustrated the potential of integrating domain-specific knowledge with intelligent decision-support mechanisms.

Although considerable research has been conducted in the individual areas of Retrieval-Augmented Generation, Agentic AI, Multi-Agent Systems, and computer-vision-based agricultural applications, their combined potential offers further opportunities for exploration. In particular, integrating computer vision with a multi-agent RAG framework could provide an intelligent approach to automated grain quality assessment. Such a framework could combine visual impurity detection and quantification with the retrieval of relevant quality standards, collaborative reasoning, and evidence-supported decision-making. This presents a promising direction for developing intelligent and knowledge-grounded grain quality assessment systems.

3. PROPOSED MULTI-AGENT RAG FRAMEWORK FOR GRAIN QUALITY ASSESSMENT

The proposed framework integrates computer vision, Retrieval-Augmented Generation (RAG), and Agentic AI to facilitate automated grain quality assessment. It consists of six specialized AI agents that collaboratively process grain images, quantify detected impurities, retrieve relevant quality specifications, perform quality reasoning, make assessment decisions, and generate a structured quality report.

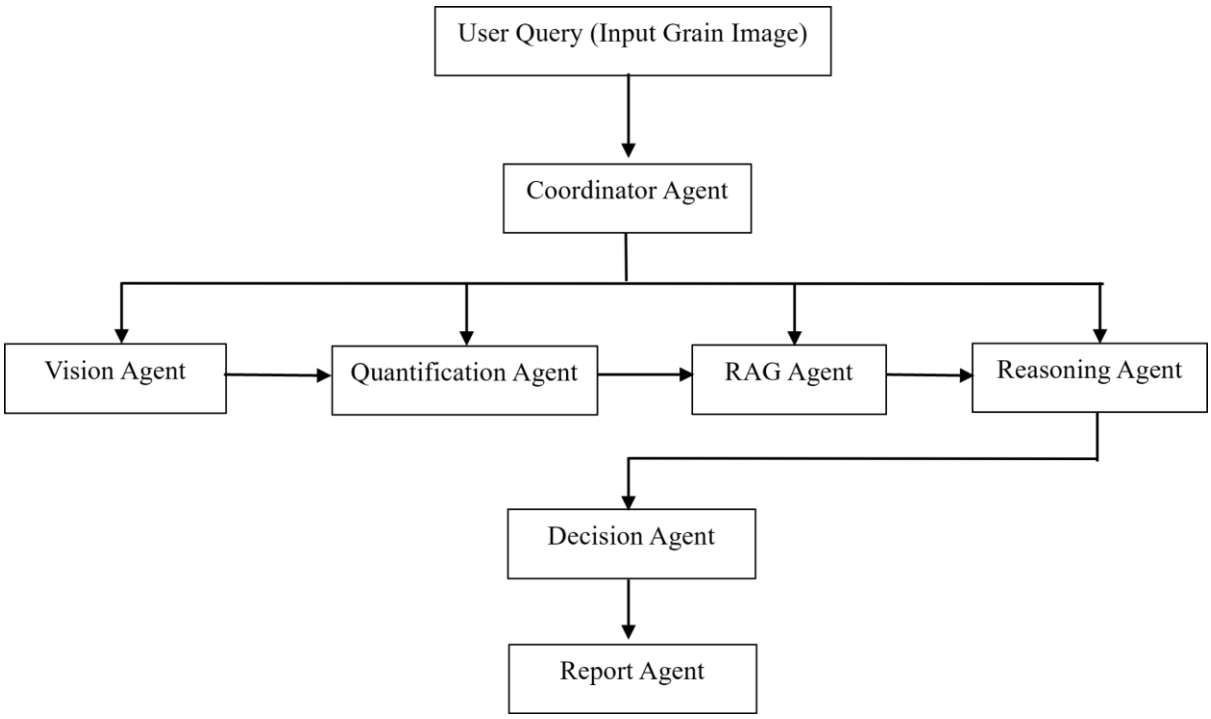

Fig 1. Multi-Agent RAG framework for Grain Quality Assessment

Fig. 1 represents the basic working of the multi-agent RAG framework for automated grain quality assessment. The proposed framework has 7 agents that communicate and coordinate with each other to solve the given problem. The process begins when a grain sample image is provided to the system along with its grain category. The coordinator agent takes the user request/input image. It coordinates the work of all the agents. The Vision Agent identifies the impurities present in the image using a trained computer vision model. The Quantification Agent subsequently calculates the number and percentage of detected impurities. These results are forwarded to the RAG Agent, which retrieves relevant grain-specific quality standards and technical guidelines from an external knowledge base. The Quality Reasoning Agent interprets the quantitative results in the context of the retrieved information. Based on this interpretation, the Decision Agent determines the appropriate quality assessment outcome. Finally, the Report Agent generates a structured report containing the detected impurities, their quantities, relevant quality criteria, and the final assessment.

4. MATHEMATICAL FORMULATION OF THE PROPOSED FRAMEWORK

This includes the operations performed by the individual agents using a set of mathematical formulations. The proposed framework integrates computer vision-based impurity detection, quantitative analysis, knowledge retrieval, quality reasoning, decision-making, and report generation. Each agent processes the information received from the preceding stage and generates structured outputs that contribute to the overall grain quality assessment. The mathematical formulations describe the major operations involved in the framework, including the identification and quantification of impurities, retrieval of relevant grain quality specifications. These formulations provide a structured representation of the collaborative workflow and explain how visual information and domain-specific knowledge are integrated. The vision agent identifies impurities in the input grain image as shown in equation 1.

$$D = f_\theta(I) \qquad (1)$$

Here, I represent the input grain image, f_θ represents the trained object detection or instance segmentation model, and D represents the set of detected impurities. The quantification agent calculates the percentage of detected impurity present in the given grain image as shown in equation 2.

$$P_i = \frac{n_i}{N_{total}} \times 100 \qquad (2)$$

The P_i represents the percentage of detection impurity. n_i represents the number of detected impurities, and N_{total} represents the total number of grains present in the given input image.

$$Sim(q, d) = \frac{E_q \cdot E_d}{||E_q||\ ||E_d||} \quad (3)$$

The RAG Agent retrieves relevant grain-quality standards from its knowledge base as shown in equation 3. Here, q represents user query or grain-quality assessment query. d represents candidate document or the document chunk. $E_q$is the query embedding, and E_d is the document embedding.

The quality reasoning agent compares the quantified impurity measurements with the retrieved quality standards. For example, if a particular standard specifies a maximum permissible impurity percentage, the compliance condition can be represented as:

$$C_i = \begin{cases} 1, & P_i \leq T_i \\ 0, & P_i > T_i \end{cases} \quad (4)$$

Here, P_i = is the measured impurity percentage, T_i is the maximum permissible limit retrieved from the applicable standard, and C_i is the compliance indicator.

5. CONCLUSION

The integration of Generative Artificial Intelligence, Retrieval-Augmented Generation, and Agentic AI has enabled decision support systems to work on complex problems and make concrete decisions. This paper presented a survey of existing research contributions in the areas of RAG, autonomous AI agents, Multi-Agent Systems, Multi-Agent RAG, and computer-vision-based agricultural applications. The reviewed studies highlighted the significance of the integration of the RAG framework and Agentic AI for today's complex tasks. Furthermore, this paper proposed a conceptual seven-agent Multi-Agent RAG framework for automated grain quality assessment. The framework integrates computer vision-based impurity detection and quantification with external knowledge retrieval, quality reasoning, decision-making, and structured report generation. This all is controlled by the coordinator agent.

In future work, the proposed framework can be implemented and evaluated using real-world grain image datasets and authoritative grain quality standards. Further research may focus on improving inter-agent communication, retrieval accuracy, reasoning consistency, and the reliability of quality assessment decisions. The framework can also be extended to support additional grain categories, more comprehensive impurity measurements, and human-assisted validation. Such developments could contribute towards building intelligent, transparent, and automated grain quality assessment systems.

REFERENCES

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in Advances in Neural Information Processing Systems (NeurIPS), vol. 27, pp. 2672–2680, 2014, https://doi.org/10.48550/arXiv.1406.2661

[2] Ho, A. N. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 6840–6851, 2020, https://doi.org/10.48550/arXiv.2006.11239

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017, https://doi.org/10.48550/arXiv.1706.03762

[4] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, 2020, https://arxiv.org/abs/2005.11401

[5]. S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. R. Narasimhan, and Y. Cao, "ReAct: Synergizing Reasoning and Acting in Language Models," in International Conference on Learning Representations (ICLR), 2023, https://arxiv.org/abs/2210.03629

[6]. L. Wang et al., "A Survey on Large Language Model Based Autonomous Agents," Frontiers of Computer Science, vol. 18, Art. no. 186345, 2024, https://doi.org/10.1007/s11704-024-40231-1.

[7]. X. Li, S. Wang, S. Zeng, Y. Wu, and Y. Yang, "A Survey on LLM-Based Multi-Agent Systems: Workflow, Infrastructure, and Challenges," Vicinagearth, vol. 1, Art. no. 9, 2024, https://doi.org/10.1007/s44336-024-00009-2.

[8]. T. Guo, X. Chen, Y. Wang, R. Chang, S. Pei, N. V. Chawla, O. Wiest, and X. Zhang, "Large Language Model Based Multi-Agents: A Survey of Progress and Challenges," in Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI), Survey Track, pp. 8048–8057, 2024, https://doi.org/10.24963/ijcai.2024/890.

[9]. T. Nguyen, P. Chin, and Y.-W. Tai, "MA-RAG: Multi-Agent Retrieval-Augmented Generation via Collaborative Chain-of-Thought Reasoning," arXiv preprint arXiv:2505.20096, 2025, https://arxiv.org/abs/2505.20096.

[10]. C.-Y. Chang, Z. Jiang, V. Rakesh, M. Pan, C.-C. M. Yeh, G. Wang, M. Hu, Z. Xu, Y. Zheng, M. Das, and N. Zou, "MAIN-RAG: Multi-Agent Filtering Retrieval-Augmented Generation," in Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL), pp. 2607–2622, 2025, https://doi.org/10.18653/v1/2025.acl-long.131.

[11]. A. Singh, A. Ehtesham, S. Kumar, T. T. Khoei, and A. V. Vasilakos, "Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG," arXiv preprint arXiv:2501.09136, 2025, https://doi.org/10.48550/arXiv.2501.09136.

[12]. M. M. Abootorabi, A. Zobeiri, M. Dehghani, M. Mohammadkhani, B. Mohammadi, O. Ghahroodi, M. S. Baghshah, and E. Asgari, "Ask in Any Modality: A Comprehensive Survey on Multimodal Retrieval-Augmented Generation," in Findings of the Association for Computational Linguistics: ACL 2025, pp. 16776–16809, 2025, https://doi.org/10.18653/v1/2025.findings-acl.861.

[13]. H. O. Velesaca, P. L. Suárez, R. Mira, and A. D. Sappa, "Computer Vision Based Food Grain Classification: A Comprehensive Survey," Computers and Electronics in Agriculture, vol. 187, Art. no. 106287, 2021, doi: 10.1016/j.compag.2021.106287, https://doi.org/10.1016/j.compag.2021.106287.

[14]. X. Zhang, D. Li, L. Wang, and M. Wu, "An Overview of the Current Situation and Future Development Direction of Grain Detection: Taking Computer Vision Combined with Deep Learning," Journal of Food Science and Technology, vol. 62, no. 11, pp. 2013–2030, 2025, https://doi.org/10.1007/s13197-025-06428-3.

[15]. P. Zapotoczny and P. K. Graczyk, "Using Deep Learning Algorithms and a Flatbed Scanner to Assess Rice Quality," Technical Sciences, vol. 28, pp. 85–102, 2025, https://doi.org/10.31648/ts.11180.

[16]. L. Yang, X. Hu, C. Shu, J. Wang, N. Wei, T. Liu, Q. Liu, M. Zhao et al., "Machine Vision-Based Impurity Content Assessment of Paddy Grains for Pre-Storage Quality Monitoring," Journal of Stored Products Research, vol. 119, Art. no. 103158, 2026, https://doi.org/10.1016/j.jspr.2026.103158.

[17] D. F. Kandamali, W. M. Porter, E. Porter, A. McLemore, and G. C. Rains, "CottonBot: An AI-Driven Cotton Farming Assistant and Irrigation Advisor Using LLM-RAG and Agentic AI Tools," Smart Agricultural Technology, vol. 12, Art. no. 101640, 2025, https://doi.org/10.1016/j.atech.2025.101640.

[18]. D. J. S. Ravindran, I. Skarga-Bandurova, S. V., M. Awais, and M. S., "AgroLLM: Connecting Farmers and Agricultural Practices Through Large Language Models for Enhanced Knowledge Transfer and Practical Application," AgriEngineering, vol. 8, no. 1, Art. no. 38, 2026, https://doi.org/10.3390/agriengineering8010038