Multimodal Data Fusion in Smart Agriculture: Integrating Soil, Climate, Imaging, and Biological Signals Using Machine Learning: Scoping Review

R.N.I. Basnayake*

Department of Computer and Data Science, NSBM Green University, Sri Lanka

DOI: https://doi.org/10.51584/IJRIAS.2026.111500010

Received: 16 April 2026; Accepted: 22 April 2026; Published: 14 May 2026

ABSTRACT

The evolution of smart agriculture has expedited the adoption of machine learning-based solutions for crop monitoring, disease detection, yield estimation, and environmental management. However, most of the prevailing machine learning-based solutions utilize unimodal data, such as image, soil, or climatic-based information. These solutions face the limitation of effectively handling the complex interactions between the agricultural ecosystems. Multimodal data fusion-based solutions have been recognized as a powerful tool for effectively utilizing the strength of heterogeneous data sources, including soil-based information, climatic-based information, image-based information, and plant physiological-based information. This systematic review focuses on the machine learning-based multimodal data fusion solutions developed using research published between 2020 and 2026. A structured review-based methodology was followed to identify the relevant research. A review of the relevant literature was conducted to determine the effectiveness of the machine learning-based multimodal data fusion solutions. It was found that the effectiveness of the machine learning-based solutions, including early, intermediate, and late fusion-based solutions, attention-based solutions, transformer-based solutions, graph learning-based solutions, and deep ensemble-based solutions, was better compared to unimodal-based solutions. The applications of the machine learning-based multimodal data fusion solutions were found to be effective for applications such as crop disease detection, soil moisture estimation, yield estimation, plant stress detection, and intelligent decision-making. The challenges associated with the machine learning-based multimodal data fusion solutions were identified to be associated with issues such as heterogeneity, synchronization, the unavailability of standardized multimodal datasets, and the limitations associated with real-time applications. The research direction of the machine learning-based multimodal data fusion solutions was identified to be associated with cross-modal transformer-based solutions, explainable machine learning-based solutions, federated learning-based solutions, the utilization of biological-based information with remote sensing-based information, and the utilization of graph learning-based solutions.

Keywords: Multimodal learning; Deep learning; Information fusion; Crop monitoring; Soil–climate integration

INTRODUCTION

The increasing global food demand, in association with variable climatic conditions and resource depletion, has accelerated the development of smart and precise farming systems. The integration of artificial intelligence (AI) and machine learning (ML) has been a key feature of Agriculture 4.0, enabling automated crop, environment, and decision-making support systems [1]. Recent developments in ML-based methods have demonstrated their potential to improve disease detection, yield forecasting, soil analysis, and stress monitoring in various agricultural environments [2].

Traditionally, agricultural intelligence systems have been largely dependent on unimodal data sources, including RGB images, soil moisture, and climatic conditions. Though these methods have been found to produce impressive results, their ability to address the multidimensional nature of agricultural environments has been limited [3]. Agricultural environments are highly heterogeneous, and crop health or productivity is influenced by various soil, climatic, spectral, and physiological factors [4].

To overcome the challenges associated with unimodal analysis, a robust concept of multimodal data fusion has been proposed to effectively integrate heterogeneous agricultural data sources. Recent research efforts have demonstrated the effectiveness of multimodal fusion approaches to integrate image, sensor, and textual modalities using the transformer-based fusion mechanism to improve the accuracy of disease detection [5, 6]. In a similar direction, multimodal fusion approaches that integrate satellite image and climatic modalities have been shown to improve the performance of yield prediction and soil moisture estimation [7, 8]. Recent advancements in the development of sophisticated fusion models, including graph-based multimodal learning [9], deep ensemble-based fusion models [10], attention-based fusion models [11], further emphasize the robustness of the multimodal fusion concept.

In addition, comprehensive review articles emphasize the growing significance of information fusion in the context of smart agriculture, including the integration of remote sensing, IoT-based sensors, machine learning, etc., to ensure the sustainable management of agricultural resources [1, 12]. However, the research community lacks a systematic review of the current trends, methodologies, applications, and challenges associated with machine learning-based multimodal fusion approaches in the context of smart agriculture. Therefore, the current study provides a comprehensive review of machine learning-based multimodal fusion approaches in the context of smart agriculture, including fusion methodologies, architectural trends, applications, and research gaps from 2020 to 2026.

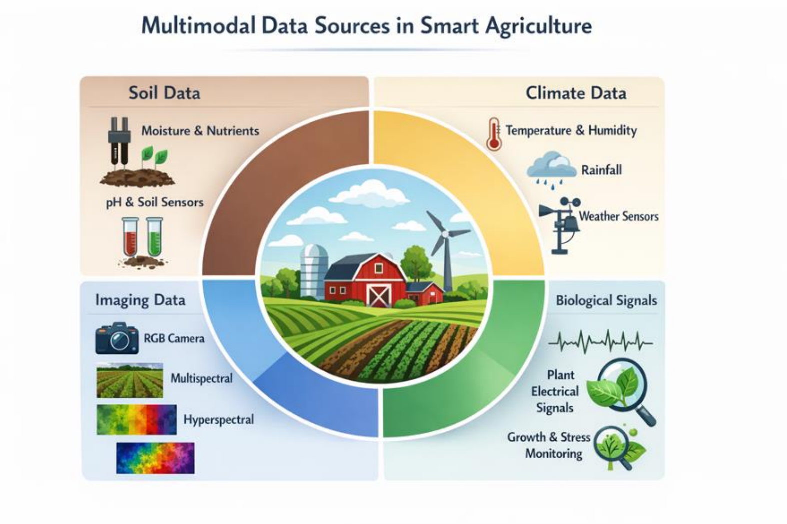

Figure 1. Key data modalities used in multimodal smart agriculture systems, including soil, climate, imaging, and biological signals.

METHODS

Review Design

This study follows a systematic literature review (SLR) research methodology to explore the recent developments in machine learning-based multimodal data fusion techniques used in smart agriculture. The use of the SLR research methodology ensures a transparent process of selecting, evaluating, and synthesizing the relevant scientific studies. The review focuses on studies published within the time frame of 2020-2026, a period when significant developments in artificial intelligence and multimodal learning techniques have been seen in the field of agricultural studies. The overall process of the review involves a series of steps, including the identification, screening, evaluation, and synthesis of the studies. The studies are used to understand the amalgamation of heterogeneous data sources in the field of agriculture, including soil, climatic, remote sensing, and physiological data.

Research Questions

The study is based on specific research questions, which guide the review and ensure a detailed analysis of the literature on the topic of multimodal data fusion in smart agriculture. The investigation covers the data modalities used in the multimodal agricultural system, which include data from the soil, climate, imaging, and biology. It also covers the assessment of the machine learning and deep learning models used for data fusion in agriculture. The review covers the major agricultural applications supported by the multimodal machine learning models, which include crop disease detection, crop yield prediction, soil monitoring, and crop stress analysis. It also covers the major challenges, limitations, and gaps in the development and implementation of multimodal agricultural intelligence systems. Therefore, the research questions guide the analysis of the selected studies and the synthesis of the results discussed in the review.

Eligibility Criteria

For ensuring the relevance and research rigor of the reviewed articles, clear inclusion and exclusion criteria have been set. The articles reviewed for the purpose of the research are limited to journal articles and conference articles published within the time frame of 2020 to 2026. The articles reviewed are those that specifically deal with the application of various multimodal data fusion methods based on machine learning or deep learning. The articles reviewed are also limited to those that specifically deal with agricultural data and those that are not based on unimodal data sources. Additionally, the articles reviewed are limited to those that are published in the English language and have accessible full-content information.

Information Sources and Search Strategy

The relevant literature has been obtained from various scholarly databases, such as Scopus, Web of Science, IEEE Xplore, and Google Scholar. The search keywords and Boolean operators were fixed to search for relevant literature. Some examples of search queries include the use of words such as ‘multimodal data fusion,’ ‘machine learning,’ ‘smart agriculture,’ ‘precision agriculture,’ ‘crop monitoring,’ and ‘remote sensing.’ These words have been combined by using logical operators such as ‘AND’ and ‘OR’ to cover the research domain completely.

Study Selection

This study used a multistep process of filtering articles that were retrieved. First, the articles were filtered through their titles and abstracts. This was done to rule out any articles that would be irrelevant to the research. Then, the relevant articles were assessed at the next level using the full texts in order to check their suitability to the topic.

Data Charting and Extraction

Data extraction was performed for each selected research article in such a manner that information necessary for analysis was collected in a systematic way. Information that was extracted included agricultural applications, data modes used, ML methods employed, fusion methods, datasets, and metrics used in each study. Through the process of information extraction, some trends, popular models, and applications of MAI were identified.

Quality Assessment / Risk of Bias

For the verification of the quality of synthesis, an analysis was made of the studies chosen based on a number of quality criteria that included methodology, explanation of the dataset, experiment design, and performance measures. The experiments carried out using clear methodologies and repeatable procedures were less likely to have any form of biases compared to those where the methodology was not well described.

RESULTS AND SYNTHESIZED FINDINGS

This analysis demonstrates the quick development in multimodal data fusion in smart agriculture. This section provides a synthesis of results in terms of critical factors such as data modalities, machine learning models, fusion approaches, and application areas.

Data Modalities in Smart Agriculture

In the analyzed papers, researchers use various data modalities due to the multidimensionality of agriculture processes. Common data sources are soil characteristics (e.g., moisture and nutrients), climate variables (temperature, humidity, rainfall), and image datasets (RGB, multispectral, hyperspectral images). Also, there are several cases where biological and physiological data (plant electricity) are used as an input to machine learning models to provide insights into current plant conditions. Thus, multimodal approach enables a more extensive representation of agricultural environments than the one that relies on a single data type.

Machine Learning and Deep Learning Models

A wide range of machine learning and deep learning models is used to solve agricultural problems. Traditional ML algorithms like Random Forest and Support Vector Machines are utilized for some tasks associated with data coming from sensors. Meanwhile, most of the contemporary research applies modern deep learning models like CNN, Vision Transformers, Graph Neural Networks (GNNs), and attention architectures. Multimodal transformers have shown good performance in detecting complex interconnections between data from various modalities. Also, there are instances where deep ensemble models are utilized to obtain better results.

Multimodal Data Fusion Strategies

Three major approaches to data fusion are discussed in the literature: early, intermediate (feature-level), and late (decision-level) fusion. While early fusion consists in combining raw data from all modalities before applying machine learning models, in intermediate fusion raw data is represented by features extracted by models. Finally, in late fusion outputs from separate models are combined. Modern fusion techniques include utilization of attention mechanisms and embeddings which enable dynamic weighing of importance of each modality. Overall, these advanced approaches allow obtaining higher accuracy of the model.

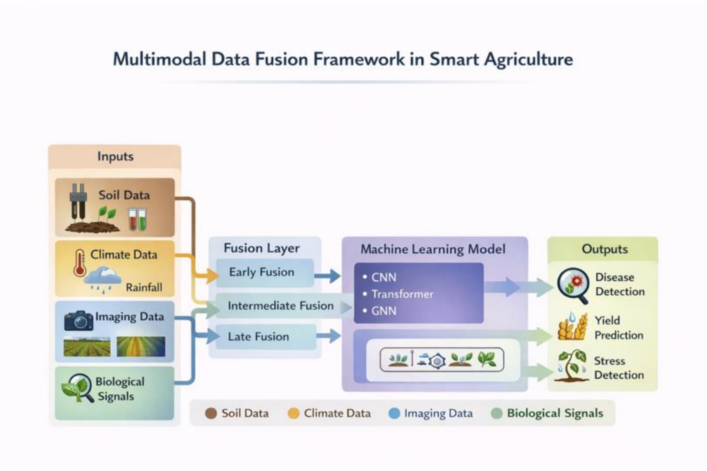

Figure 2. General framework of multimodal data fusion in smart agriculture, illustrating the integration of heterogeneous data sources, fusion strategies, and machine learning models for agricultural decision-making.

Comparative Analysis of Multimodal Fusion Approaches

Different fusion approaches have unique strengths and weaknesses, depending on the application in agriculture and the nature of data. The early fusion approach has lower computation costs but may face problems in integrating diverse data sources. Intermediate fusion allows for better feature extraction but requires more complicated model architectures. Late fusion allows independent training of each model but may not be capable of capturing deep cross-modal interdependencies.

Transformer models have shown their capability to extract cross-modal relationships, especially when it comes to detecting diseases and predicting yields. However, they are data-intensive and require high computational power. Models based on graph neural networks effectively encode spatial and relational dependencies. Nonetheless, they may be difficult to implement in real-time environments.

Models based on attention mechanisms improve the interpretability of fusion results since they assign weights to different modalities. Yet attention models can also increase computational costs. Ensemble models ensure robustness and accuracy of fusion results but take longer training periods and more computing resources. No fusion approach can claim universality. The selection of an adequate technique relies on multiple factors, such as available data, computational resources, and particular use cases.

Quantitative Performance Insights

Despite the different data sources and assessment criteria used in the studies, multimodal methods tend to outperform their unimodal counterparts. The reported accuracy increases vary between 5 and 20 percent, with the biggest improvements demonstrated in the context of crop disease identification and crop yield forecasting. Methods for combining transformers and attention-based techniques demonstrate the most significant gains in accuracy, while improvements made by machine learning algorithms tend to be more modest.

Application Areas

There are multiple possible applications of multimodal data fusion in agriculture. Crop diseases detection seems to be the most frequent area, as image and environmental data provide great insights into diseases. Other widely studied problems are prediction of crop yields, estimation of soil moisture levels, and detection of stress states of plants. Apart from that, decision support systems based on multimodal data fusion are developed for real-time management of irrigation and crop cultivation.

Key Findings and Future Directions

The use of multimodal approaches is always superior to unimodal methods in terms of accuracy, robustness, and generalizability. Multimodal fusion helps in modeling complex relationships in agricultural ecosystems through data obtained from different sources. However, future areas of interest have been highlighted in this review. These include data heterogeneity, inadequate availability of benchmark datasets, and the implementation challenge when applying in real time. Potential areas of development involve transformers, explainable AI, and bio-sensor data along with remote sensing data.

DISCUSSION

From this review, one can conclude that multimodal data fusion has become increasingly important for developing smart agriculture. Combining various data types (soil, climate, imaging, biological data), one can provide a multidimensional description of agricultural ecosystems, which is crucial considering the complexity of plant cultivation and crop production processes determined by different factors. Deep learning models (transformers, attention mechanisms, graph neural networks) have been widely used for this purpose, demonstrating their ability to identify and leverage cross-modal dependencies and significantly increase prediction performance when applied for various tasks (disease recognition, yield estimation, stress monitoring).As compared to traditional machine learning algorithms, deep learning multimodal algorithms show better robustness and generalization performance in diverse agricultural environments. However, some challenges should be considered, among which data heterogeneity, since different types of data could have different scales and temporal resolutions. Moreover, there are no publicly available multimodal datasets for agriculture. Also, applying multimodal models to actual farms is rather complicated due to hardware requirements and technical restrictions in collecting sensor data. Finally, deep learning models, especially more sophisticated ones like transformers and deep ensembles, tend to be black boxes, which is why incorporating XAI tools into the process is critical to achieve better performance and interpretability of results. In conclusion, one can state that the reviewed articles indicate that multimodal data fusion is a promising direction to go with future advances in intelligent sustainable agriculture.

Limitations And Future Work

However, there are some problems with this systematic review. To begin with, the search restricted itself to articles published in the last 2-6 years (between 2020 and 2026). This can mean that there might have been some important studies conducted before that time which may have been omitted. The review was further limited as the article only in English was used as a criterion. Furthermore, the datasets and evaluation criteria as well as model specifications differ significantly, which does not allow to compare the performance rates of these models appropriately.

The small datasets used in the literature that are domain-specific might limit to a certain extent the generalizability of the results reported in the studies. It is important to note that the standardized set of multimodal agricultural data is yet to be established and compare the models according to the diverse methodologies.

Therefore, the creation of multimodal agricultural datasets that include all the above-mentioned types of data in the future is the way to go. The other thing that the future research will involve will be how to come up with practical solutions that would make the implementation of the discussed models to the field possible. In addition, it is crucial to use methods of explainable artificial intelligence to promote user-friendly and reliable models. Likely future research directions are through cross-modal transformers, federated learning and edge AI.

CONCLUSION

The current review provides an in-depth discussion of machine learning-based methods of multimodal data fusion in smart agriculture, with the period of studies discussed within the span of 2020-2026. The results indicate that multimodal data, incorporating soil parameters, climatic variables, imaging data, and biological cues, have a considerable positive effect on the performance, robustness, and reliability of agricultural intelligence systems as compared to single-modes data modalities. The review identifies the increasing use of more sophisticated machine learning methods, in particular deep learning architectures like convolutional neural networks, transformer-based models, graph neural networks, and attention-based fusion models. Such methods allow successful cross-modal learning and enhance predictive accuracy in a variety of agricultural uses, such as crop diseases, yield, soil surveillance, and stress diagnosis of plants. Although these are being developed, various challenges still exist, such as heterogeneity of data, unavailability of standardized multimodal data, limitations to real-time implementation, and interpretability of complex models. The multimodal systems must be put into practice successfully in real-life and this is possible by addressing these issues. Altogether, multimodal data fusion is one of the directions promising developments of smart, data-driven, and sustainable agricultural systems. Subsequent studies ought to concentrate on scalable designs, better data assimilation plans, and the integration of explainable and real-time solutions and live remedies to aid feasible agricultural choices.

REFERENCES

1. Choi, J. W., Cho, S. B., Hidayat, M. S., Hwang, W.-H., Cho, Y.-S., Lee, H., Cho, B.-K., & Kim, G. (2025). Application of multimodal data fusion and explainable AI for classifying water stress in sweet potatoes. *Frontiers in Plant Science*, 16.
2. Katharria, A., Rajwar, K., Pant, M., Velásquez, J. D., Snášel, V., & Deep, K. (2025). Information fusion in smart agriculture: Machine learning applications and future research directions. *arXiv preprint arXiv:2405.17465*.
3. Khoirudin, P. T., Pungkasanti, P. T., & Hidayati, N. (2024). A systematic review on data fusion techniques for agricultural yield prediction: Integrating satellite imagery with climatic data. *Systematic Literature Review Journal*, 1(4).
4. Lamichhane, M., Mehan, S., & Mankin, K. R. (2025a). Multimodal machine learning approaches for soil moisture estimation in agricultural systems.
5. Lamichhane, M., Mehan, S., & Mankin, K. R. (2025b). Surface soil moisture prediction using multimodal remote sensing data fusion and machine learning algorithms in semi-arid agricultural regions.
6. Liu, R., Chang, C., Zhong, R., & Lu, S. (2025). Soil moisture monitoring method and data products: Current research status and future development trends. *Remote Sensing*, 17, 3945.
7. Lu, Y., Lu, X., Zheng, L., Sun, M., Chen, S., Chen, B., Wang, T., Yang, J., & Lv, C. (2024a). Application of multimodal transformer model in intelligent agricultural disease detection and question-answering systems. *Plants*, 13, 972.
8. Lu, Y., et al. (2024b). Multimodal transformer-based intelligent systems for agricultural decision support. *Plants*.

9. Njoroge, T. K., Sindu, K. M., & Kibuku, R. (2025). Edge-optimized multimodal cross-fusion model with statistical validation for multi-crop disease detection. *International Journal of Advances in Intelligent Informatics*, 11(3).
10. Shoaib, M., Khan, S. U., AbdelHameed, H., & Qahmash, A. (2026a). Plant stress detection using multimodal imaging and machine learning: From leaf spectra to smartphone applications. *Frontiers in Plant Science*.
11. Shoaib, M., et al. (2026b). Machine learning-driven multimodal imaging techniques for plant stress analysis. *Frontiers in Plant Science*.
12. Wang, X., Yan, F., Li, B., Yu, B., Zhou, X., Tang, X., Jia, T., & Lv, C. (2025). A multimodal data fusion and embedding attention mechanism-based method for eggplant disease detection. *Plants*, 14, 786.
13. Wittstruck, L., Waske, B., & Jarmer, T. (2025). Multi-modal vision transformer for high-resolution soil texture prediction using remote sensing imagery. *Remote Sensing of Environment*.
14. Wu, X., Zhang, J., Zou, Z., Chen, C., Yu, Y., Yu, P., Xiao, Y., Wang, Q., & Hao, G. (2025a). PlantIF: Multimodal semantic interactive fusion via graph learning for plant disease diagnosis.
15. Wu, X., et al. (2025b). Graph-based multimodal learning for intelligent plant disease diagnosis systems.
16. Yang, Z.-X., Li, Y., Wang, R.-F., Hu, P., & Su, W.-H. (2025a). Deep learning in multimodal fusion for sustainable plant care: A comprehensive review. *Sustainability*, 17(12), 5255.
17. Yang, Z.-X., et al. (2025b). Advances in multimodal deep learning frameworks for sustainable agriculture applications. *Sustainability*.
18. Yewle, A. D., Mirzayeva, L., & Karakuş, O. (2025). Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction. *arXiv preprint*.
19. Yeom, M.-S., Lee, Y., Oh, H., Lee, E., & Oh, M.-M. (2025). Applying machine learning for the classification of environmental conditions using plant electrical signals. *Horticulture, Environment, and Biotechnology*.
20. Zhou, Y., Yan, H., Ding, K., Cai, T., & Zhang, Y. (2024). Few-shot image classification of crop diseases based on vision–language models. *Sensors*, 24, 6109.