frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 18 June 2025
DOI 10.3389/fenvs.2025.1561655

OPEN ACCESS

EDITED BY
Ana Maria Tarquis,
Polytechnic University of Madrid, Spain

REVIEWED BY
Mohamed R. Abonazel,
Cairo University, Egypt
Ahmed Saqr,
Mansoura University, Egypt
Peng Du,
Liaoning Normal University, China

*CORRESPONDENCE
Yan Shi,
shiyan24s@sina.com

RECEIVED 20 January 2025
ACCEPTED 16 May 2025
PUBLISHED 18 June 2025
CORRECTED 09 July 2025

CITATION
Shi Y, Zhao P, Gu Z and Li Y (2025) Synergistic research on planter performance optimization and green low-carbon agricultural transformation under climate risk.
Front. Environ. Sci. 13:1561655.
doi: 10.3389/fenvs.2025.1561655

Synergistic research on planter performance optimization and green low-carbon agricultural transformation under climate risk

Yan Shi*, Pengfei Zhao, Zhengzhao Gu and Ye Li

Department of Mechanical Engineering, Taiyuan Institute of Technology, Taiyuan , China

Introduction: Addressing the dual challenges of climate change and sustainable food production, this study proposed an integrated framework that combined planter performance optimization with green, low-carbon agricultural transformation. While traditional planting strategies focused on parameters like seed depth, speed, and spacing, they often neglected environmental sustainability and adaptability to climate variability.

Methods: To bridge this gap, we introduced the Adaptive Precision Planter Optimization Model (APPOM), which leveraged real-time environmental sensing, machine learning, and multi-objective optimization to dynamically adjust key planting parameters. Our approach also incorporated green technologies, including electric-powered planters and carbon-sequestration soil practices, to reduce the ecological footprint of agricultural operations.

Results: Experimental results validated that APPOM significantly improved planting accuracy, enhanced resource efficiency, and reduced carbon emissions across diverse soil and climate conditions. Furthermore, we presented the Real-Time Adaptive Planter Optimization (RAPO) strategy, which enabled context-aware decision-making and continuous optimization under field variability.

Discussion: The findings underscored the potential of intelligent, eco-friendly planting systems to foster climate-resilient agriculture. However, challenges such as cost barriers and deployment scalability remained. Future research should aim to enhance affordability and accessibility, particularly for smallholder farmers, and expand the framework to a broader range of crops and regions.

KEYWORDS

climate resilience, precision agriculture, low-carbon farming, planter optimization, sustainability

1 Introduction

Planters are agricultural machines designed to place seeds into the soil at controlled depth, spacing, and rate, playing a critical role in ensuring seed germination and uniform crop establishment (Hu et al., 2023). Planter optimization refers to the process of adjusting these operational parameters to improve efficiency, minimize input waste, and maximize yield. Conventional planter optimization typically focuses on mechanical improvements—such as refining seed metering systems, regulating ground speed, and maintaining consistent spacing between seeds (Peng et al., 2022). These strategies often rely

on fixed rules and assume uniform field conditions, which limits their adaptability to environmental variation (Wei et al., 2023).

However, traditional approaches have notable limitations. They fail to account for spatial and temporal variability in soil properties, weather conditions, and terrain features (Zong et al., 2023). As a result, even minor deviations in field conditions—such as changes in moisture or compaction—can lead to uneven seed placement (Zhou H.-Y. et al., 2023), reduced emergence, and yield variability (Song et al., 2023). Moreover, conventional systems generally overlook sustainability factors such as energy efficiency or emissions, making them less suited for climate-resilient agriculture (Xu et al., 2022).

Despite recent progress in precision agriculture and remote sensing-based decision systems, existing studies still face several critical limitations that hinder their real-world applicability. First, most prior methods rely heavily on static optimization schemes that fail to account for dynamic environmental changes during planting, such as real-time soil moisture or compaction variability. This leads to suboptimal seed placement and inefficient resource use under heterogeneous field conditions. Although multimodal deep learning has been applied to tasks like crop classification or land cover segmentation, few works have successfully bridged low-level perception with high-level planting strategy optimization. The models often excel in classification metrics but lack interpretable connections to agronomic decision variables such as seed depth, spacing, or planting speed. Moreover, most studies utilize synthetic or benchmark datasets that do not reflect the complexity, noise, or operational constraints encountered in mechanized field deployment. Real-time feedback and adaptive control are rarely integrated into existing frameworks. Systems are typically designed to generate pre-season recommendations, without the capacity to respond to on-the-fly soil condition shifts or machinery behavior. As a result, the temporal mismatch between sensing and acting weakens their practical deployment value. These limitations motivate our design of the APPOM and RAPO frameworks, which jointly address prediction, optimization, and adaptive real-time decision-making for precision planting under diverse agricultural conditions.

Recent advances in precision agriculture offer a promising alternative (Lian et al., 2022). Through the use of real-time sensors, satellite data, GPS positioning, and machine learning algorithms, planting operations can be adapted dynamically based on environmental feedback (Yao et al., 2023). These data-driven technologies allow for more context-aware decisions, such as adjusting planting depth in response to moisture levels or altering speed to reduce soil disturbance (Zhang et al., 2023). At the same time, green technologies—such as electric-powered planters, low-emission actuators, and carbon-sequestration soil practices—are being increasingly integrated into agricultural equipment to reduce carbon footprints (Joseph et al., 2023). Together, these tools enable a more holistic approach to planter optimization, balancing productivity with sustainability goals (Zhang et al., 2022).

Given these developments, there is a growing need for integrated frameworks that combine adaptive intelligence with green transformation. This paper addresses that need by proposing a novel optimization approach that unifies advanced planter control with real-time environmental awareness and low-carbon practices. We develop two complementary systems: the Adaptive Precision Planter Optimization Model (APPOM) and the Real-Time Adaptive Planter Optimization (RAPO) strategy. These systems aim to dynamically optimize seed depth, spacing, and rate based on environmental feedback, while also reducing emissions and enhancing operational efficiency. The proposed method has several key advantages:

- Our approach integrates advanced planter optimization with green, low-carbon technologies, creating a synergistic model that addresses both performance and sustainability.
- This methodology is versatile, adaptable to various crops, regions, and climate conditions, making it an efficient and scalable solution for diverse agricultural systems.
- Our experiments demonstrate improved efficiency and reduced carbon emissions, validating the effectiveness of our approach in achieving both higher productivity and sustainability.

2 Related work

Optimizing planter performance is a crucial aspect of improving agricultural productivity (Du et al., 2022). Planters, as integral pieces of equipment in modern agriculture, play a vital role in ensuring efficient planting operations, including seed spacing, depth control, and seed-soil contact (Ren et al., 2024c). Enhancing planter performance involves adjusting parameters such as seed rate, uniformity, and germination potential to maximize yield and minimize resource waste (Li et al., 2020). These factors are especially critical under climate risk scenarios, where erratic weather conditions like droughts or floods can significantly impact planting outcomes (Lin et al., 2023).

In recent years, precision farming technologies such as GPS, sensors, and data analytics have been integrated into planting systems to allow for real-time monitoring and control (Zhou Y. et al., 2023). Variable rate planting (VRP), for instance, enables farmers to tailor seed densities to soil fertility or moisture levels (Steyaert et al., 2023). Advanced mechanical planters now include adjustable depth controllers and low-compaction designs to improve adaptability across field conditions (Ren et al., 2024b). Furthermore, some systems now incorporate weather forecasts and soil data to optimize planting schedules for climate resilience (Adeel et al., 2022).

Green low-carbon agricultural transformation has become a key objective in global climate mitigation efforts (Yan et al., 2022). Agriculture is both a contributor to and a victim of climate change, which drives the need for sustainable farming practices (Fan et al., 2022). Agroecological practices such as crop rotation, agroforestry, and cover cropping have shown great promise in reducing carbon emissions and improving soil carbon sequestration (Chango et al., 2022). These methods support long-term soil fertility and reduce environmental harm while maintaining food security Ren et al. (2024a).

Energy use in agriculture has also seen a shift toward low-emission solutions, including solar-powered irrigation and electric tractors (Taylor et al., 2018). The replacement of synthetic inputs with bio-based alternatives helps reduce emissions further (Yu et al., 2023). In addition, technologies such as biogas production and hydrogen-fueled machinery have gained traction as decarbonization tools (Wan et al., 2022). Precision agriculture,

supported by AI and IoT, also facilitates more efficient input management by helping farmers make real-time, data-informed decisions (Ektefaie et al., 2022).

To enhance resilience against climate variability, researchers have developed genetically modified crop varieties capable of withstanding environmental stressors such as drought or heat (Awwad Al-Shammari et al., 2022). Biotechnology has been instrumental in stabilizing crop yields under volatile weather patterns (Wu et al., 2022). In parallel, resource-efficient techniques like drip irrigation and rainwater harvesting address water scarcity concerns (Chai and Wang, 2022). Diversified cropping systems and practices such as agroforestry also enhance biodiversity and reduce vulnerability to pest outbreaks (Yang et al., 2022). Soil conservation strategies further contribute to long-term agricultural resilience by protecting critical ecosystem functions (Smith et al., 2021).

The integration of precision planting systems with climate risk data presents a holistic approach to sustainable agriculture (Bayoudh et al., 2021). For example, systems that optimize seeding depth and scheduling based on environmental feedback can reduce both input costs and ecological impacts (Adeel et al., 2022). Planter performance optimization referred to the set of techniques and strategies designed to enhance the efficiency, accuracy, and overall effectiveness of automated planting systems, commonly used in agriculture and horticulture. These systems, often consisting of robotic planters, precision agriculture technologies, and IoT-based solutions, were increasingly being employed to meet the growing global demand for food while addressing sustainability concerns. Optimizing planter performance was critical for improving crop yields, reducing operational costs, and ensuring environmentally sustainable practices in modern farming.

Recent developments in machine learning and multimodal data fusion have also influenced agricultural optimization. One study proposed a federated learning approach that integrates multiple agricultural data sources while preserving data privacy across farms (Cheng et al., 2025). Another proposed a multimodal learning architecture to improve field-level decisions by combining satellite imagery, sensor streams, and management logs (Jiang et al., 2023). Others showed that environmental data and sensor fusion could be used to drive low-carbon farming transitions through adaptive learning frameworks Chen et al. (2024). Building on these foundations, our study introduces APPOM and RAPO, which apply real-time sensor feedback and multi-objective optimization for agricultural planter decision-making (Ma et al., 2022).

From an agronomic standpoint, researchers have shown that small variations in seeding depth can significantly impact germination and early growth, especially under moisture stress (Li et al., 2020). Seed placement accuracy has also been linked to yield gains in conservation tillage systems (Taylor et al., 2018). Soil health improvements through organic inputs and microbial management play a crucial supporting role in optimizing planter performance (Han et al., 2024). Taken together, these agronomic insights provide a complementary foundation for the technological advancements presented in this paper (Smith et al., 2021).

3 Methods

3.1 Overview

This paper introduces a novel approach to planter performance optimization. It integrates data-driven methodologies, machine learning, and real-time environmental monitoring. The goal was to create a robust framework capable of adapting to various agricultural conditions, minimizing human intervention, and maximizing operational efficiency across different planting tasks. The structure of this approach was organized into several key components:

In Section 3.2, the foundational concepts and mathematical models underlying the optimization problem were presented. This included the formalization of key performance metrics such as planting accuracy, speed, and resource utilization, as well as the constraints imposed by soil conditions, crop types, and environmental variables. Core assumptions and the data sources used to guide the optimization process were also introduced.

In Section 3.3, a new optimization model was proposed that incorporated real-time data from various sensors, such as soil moisture levels, GPS positioning, and climate data. Planting parameters such as seed depth, spacing, and planting speed were dynamically adjusted to ensure optimal performance under changing conditions. We applied advanced machine learning to predict and correct planting errors. This enabled the planter system to learn from past operations and adapt to new scenarios.

In Section 3.4, a novel strategy for optimizing planter performance was presented. The strategy merges traditional agronomic knowledge with modern computational methods. It aims to identify optimal planting practices for diverse crop types and locations.

To ensure the robustness of the optimization process, we explicitly define the parameters used in the model. The planting parameters include: seed depth (S_1), which affects germination rate and early root development; seed spacing (S_2), influencing intra-crop competition and yield density; and planting speed (S_3), which impacts placement accuracy and operational efficiency. The soil variables considered include: soil moisture (z_1), crucial for seed hydration; soil temperature (z_2), affecting germination timing; soil pH (z_3), influencing nutrient availability; and soil compaction (z_4), which can hinder root penetration and seedling emergence. The environmental variables comprise: rainfall levels (z_5), wind speed (z_6), and solar radiation or sunlight intensity (z_7), all of which contribute to field conditions at seeding time. These factors were selected due to their direct agronomic impact and sensitivity to climate variability. For example, erratic rainfall patterns affect optimal seeding time and moisture availability; rising temperatures alter the soil thermal profile, influencing depth calibration; and higher frequency of extreme weather conditions (e.g., windstorms or droughts) necessitates dynamic adjustment of planting speed and spacing. By including these variables, our model captures a comprehensive view of field dynamics and enables adaptive decision-making that is resilient to both seasonal shifts and long-term climate change.

3.2 Preliminaries

In this section, we formalized the problem of planter performance optimization and introduced the necessary

mathematical models, assumptions, and data structures that served as the foundation for our proposed optimization approach. Consider a planter system designed to plant seeds at predetermined depths and spacings in a field. Let $S \in \mathbb{R}^{n\times m}$ represent the set of planting parameters, where n denoted the number of planting rows and m represented the number of planting positions in each row. Each element S_{ij} corresponded to the planting parameter for the j-th position in the i-th row, which could include attributes such as seed depth, spacing, or planting speed. We denoted the optimal set of planting parameters as S^*, which minimized the overall planting error and maximized planting efficiency. The performance of the planter system was influenced by various factors, including soil conditions, seed variety, environmental variables, and the mechanical capabilities of the planter. To formalize this, we defined the performance function $P(S)$, which quantified the effectiveness of the planter system based on a set of performance metrics. These metrics included ensuring that each seed was placed at the optimal depth and spacing for germination, the speed at which the planter operated while maintaining required accuracy, and the efficient use of resources such as fuel, seed, and labor. We assumed that planting performance could be expressed as a function of S, the planting parameters, and a set of external variables such as soil properties, weather conditions, and crop type. The overall objective of optimization was to minimize the following loss function (Equation 1):

$$\mathcal{L}(S) = \alpha_1 \mathcal{L}_{\text{accuracy}}(S) + \alpha_2 \mathcal{L}_{\text{speed}}(S) + \alpha_3 \mathcal{L}_{\text{resource}}(S), \tag{1}$$

where $\mathcal{L}_{\text{accuracy}}(S)$, $\mathcal{L}_{\text{speed}}(S)$, and $\mathcal{L}_{\text{resource}}(S)$ represented the error terms associated with accuracy, speed, and resource utilization, respectively. The weights α_1, α_2, and α_3 were hyperparameters that controlled the trade-off between these objectives, and were determined based on specific farming goals, such as prioritizing speed over accuracy in certain conditions.

The planting parameters S were subject to several constraints imposed by the agricultural environment, mechanical limitations, and agronomic best practices. These constraints could be divided into two categories. The first category ensured the planter system operated within its mechanical capabilities. For instance, the allowable seed depth at any position j in row i was constrained as Equation 2:

$$d_{\min} \le S_{ij} \le d_{\max}, \tag{2}$$

where $d_{\min}$ and $d_{\max}$ represented the minimum and maximum allowable values for seed depth, respectively. The second category of constraints was based on agronomic principles that dictated optimal planting conditions for different crops. For example, the set of optimal planting parameters for the j-th position in row i, based on the specific soil condition x_i and crop type y_j, was represented as Equation 3:

$$S_{ij} \in \mathcal{S}_{\text{optimal}}(x_i, y_j), \tag{3}$$

which ensured that the system adhered to agronomic guidelines for seed depth, spacing, and other planting parameters that influenced seed germination and crop growth. The performance of the planter system was also influenced by dynamic environmental factors such as weather conditions and soil properties. These factors were modeled as time-varying variables, which introduced uncertainty into the optimization process. Let $\mathbf{z}(t)$ represent the vector of environmental variables at time t, which may included factors like soil moisture $z_1(t)$, soil temperature $z_2(t)$, and weather conditions $z_3(t)$, such as rainfall and wind speed. To optimize the planter's performance in real time, the system had to adapt to these changing conditions. We modeled the influence of these environmental factors on the planting parameters as follows (Equation 4):

$$S(t) = \mathcal{G}(S, \mathbf{z}(t)), \tag{4}$$

where $\mathcal{G}$ was a function that dynamically adjusted the planting parameters S based on real-time environmental conditions. For example, the system might adjust seed depth or spacing in response to variations in soil moisture or temperature to ensure optimal seed germination. To optimize the planter's performance in real time, we relied on multiple data sources, including sensors to measure parameters such as soil moisture and temperature at different locations in the field, monitoring weather conditions like temperature, humidity, and wind speed, and tracking the location of the planter to ensure precise seed placement. These sensors generated large amounts of data, which were processed using machine learning algorithms to optimize the planting parameters over time. - $S \in \mathbb{R}^{n\times m}$: Planting parameters, where n was the number of rows and m was the number of positions per row. - $\mathcal{L}(S)$: The total loss function to minimize, combining accuracy, speed, and resource utilization. - $\mathcal{S}_{\text{optimal}}(x_i, y_j)$: The optimal set of planting parameters based on specific crop and soil conditions. - $\mathbf{z}(t)$: The vector of environmental variables at time t, including soil moisture, temperature, and weather conditions. - $\mathcal{G}(S, \mathbf{z}(t))$: A function that adjusted the planting parameters based on environmental data.

3.3 New model: adaptive precision planter optimization model (APPOM)

In this section, we presented the Adaptive Precision Planter Optimization Model (APPOM), a novel framework designed to optimize planter performance in dynamic agricultural environments. APPOM integrated machine learning, real-time environmental feedback, and multi-objective optimization to dynamically adjust planting parameters, ensuring precise seed placement and efficient resource utilization. The following sections highlighted the three core innovations of APPOM: dynamic parameter adjustment, real-time environmental integration, and adaptive learning for optimization (As shown in Figure 1). To support multimodal decision-making in agricultural contexts, our system utilizes a modified CLIP encoder to embed both visual and contextual textual inputs into a unified semantic space. In our setting, the visual encoder processes multispectral or remote sensing images (e.g., Sentinel-2, UAV imagery), while the text encoder handles structured environmental labels such as soil type, moisture class, or operational instructions. By jointly embedding these modalities, the model can interpret complex field conditions and adjust planting decisions accordingly. Attention maps derived from modules such as DINO and SAM are used to localize agronomically relevant features—such as heterogeneous soil zones, vegetation health patterns, or areas prone to waterlogging—within the input imagery. These maps inform the adaptive adjustment of planting parameters by directing focus to high-impact regions, thereby enhancing the spatial precision and environmental relevance of planting strategies.

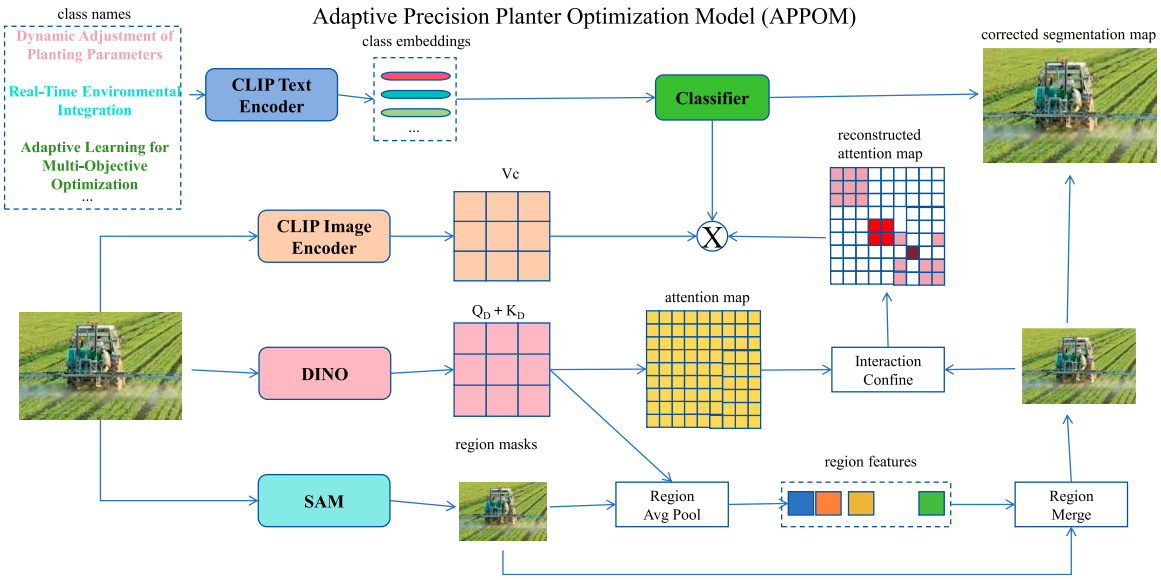

FIGURE 1
Adaptive Precision Planter Optimization Model (APPOM) Framework. This diagram illustrated the structure of the APPOM framework, which integrated multiple components for real-time adaptive planting optimization. The system utilized a CLIP-based text and image encoder to process environmental data, such as soil moisture, temperature, and other field conditions. The figure highlighted key modules including DINO, SAM, and attention mechanisms that facilitated dynamic adjustments in planting parameters like seed depth, spacing, and speed. The framework incorporated region-based attention maps for efficient resource utilization and ensured optimal seed placement by adapting to changing environmental feedback. The final output included a corrected segmentation map reflecting the optimized planting strategy.

The translation of SAR or optical image features from non-agricultural domains (e.g., ship or flood detection) into agricultural planting decisions is enabled by the structural similarity in spatial analysis tasks across these domains. Both domains require the extraction of regional contrasts, boundary contours, and context-aware object localization from remotely sensed imagery. In ship detection, for instance, the model learns to identify high-salience regions under radar speckle noise, while in agriculture, analogous attention must be given to heterogeneous soil zones, moisture-retaining depressions, or compacted strips within a field. Within the APPOM framework, this transfer is operationalized by the use of cross-modal embedding via CLIP encoders, where SAR/optical imagery is aligned with agronomic semantic tags (e.g., loamy, dry, compacted, shaded). Attention maps generated through DINO and SAM localize regions with unique spectral or textural signatures—such as rough soil patches or high-reflectance zones—which are then interpreted as indicators of environmental variability. These spatial cues do not directly output seed depth or soil compaction values but serve as proxies to modulate the downstream predictive functions. When fused with *in situ* sensor data (e.g., real-time soil moisture, temperature, or pH), these visual embeddings contribute to a multimodal decision space where the model predicts optimal planting parameters like depth or spacing. For example, a region identified as high-reflectance and low-texture in SAR imagery may be cross-referenced with sensor-indicated dryness, leading the model to increase seed depth accordingly. In this way, SAR-derived features support contextual differentiation of planting zones and enhance the granularity and precision of real-time seeding decisions.

3.3.1 Dynamic adjustment of planting parameters

A key innovation of APPOM was its ability to dynamically adjust planting parameters $S = \{S_1, S_2, \ldots, S_m\}$ in response to real-time environmental variations and performance feedback. In real-world farming, even within the same field, microzones of variability such as uneven soil compaction or inconsistent moisture levels can significantly impact planting success. Traditional planters operate with fixed parameters, which may lead to underperformance in these microzones. APPOM addresses this issue by adaptively fine-tuning parameters like seed depth and spacing based on real-time data, enhancing planting uniformity and yield outcomes. These parameters included seed depth (S_1), seed spacing (S_2), and planting speed (S_3), which were critical for ensuring optimal seed placement and maximizing crop yield. The adjustments were governed by a learned mapping function $\mathcal{A}(S, \mathbf{z}(t))$, where $\mathbf{z}(t) = [z_1(t), z_2(t), \ldots, z_k(t)]$ was a vector of real-time environmental variables at time t. These variables included critical field data such as soil moisture, soil temperature, rainfall levels, wind speed, and soil compaction. Machine learning models, including neural networks and gradient-boosted regression models, were used to approximate the function $\mathcal{A}$. These models were trained on historical planting data that captured the relationships between environmental conditions $\mathbf{z}(t)$, planting parameters S, and resulting crop yields. The historical planting data used to train the optimization model were obtained from a combination of publicly available agricultural datasets, long-term sensor logs, and field experiment records collected by agronomic institutions and research trials. These datasets contained time-series records of planting operations, environmental measurements, and yield outcomes across different crop types and regions. To account for crop-specific variations, each training instance was annotated with

metadata including crop species, cultivar, and regional soil profiles. The model was trained using a stratified sampling approach to ensure that representative examples for each crop category were included. Furthermore, crop-specific agronomic constraints—such as optimal seed depth and spacing ranges—were incorporated into the model via the constraint set $\mathcal{S}_{\text{optimal}}(x_i, y_j)$. This allowed the optimization process to remain biologically and operationally valid across diverse planting scenarios. The inclusion of such structured metadata enabled the CMDN framework to generalize across crop types while retaining the ability to fine-tune recommendations for specific planting contexts.

During deployment, the system used the learned $\mathcal{A}$ to predict and adjust the optimal planting parameters $S(t)$ in real time. In intuitive terms, this function tells the system: "if the soil is dry and compact now, increase seed depth slightly and slow down planter speed to ensure better contact and germination." The adjustment mechanism was iterative and feedback-driven, incorporating both real-time sensor data and historical performance data to continuously refine its predictions. At each time step t, the system evaluated the deviation between the current planting configuration $S(t)$ and the optimal configuration $S^{\text{optimal}}(t)$. The goal was to minimize this deviation through the following adjustment loss function (Equation 5):

$$\mathcal{L}_{\text{adjust}}(S) = \sum_{i=1}^{m} \|S_i^{\text{optimal}}(t) - S_i(t)\|^2 + \lambda \sum_{j=1}^{k} \|z_j(t) - z_j^{\text{target}}\|^2, \quad (5)$$

where $S_i^{\text{optimal}}(t)$ was the ideal value of parameter S_i, z_j^{target} represented the target environmental condition for variable z_j, and λ was a regularization factor that balanced parameter adjustments with environmental constraints. The first term penalized deviations from agronomically optimal planting configurations under the current field conditions, while the second term ensured that decisions did not violate important environmental thresholds. For example, if soil compaction exceeded normal levels, the model might reduce planting speed or increase spacing to mitigate excessive pressure on the seedbed. To further enhance its robustness, APPOM integrated predictive models that accounted for temporal variations in environmental conditions. For instance, if rainfall was forecasted, the model predicted the upcoming impact on soil moisture and proactively adjusted seed depth in advance to avoid seed oversaturation or floating. While the adjustment of seed depth based on approaching rainfall may not be universally necessary, it is particularly beneficial under highly variable climate scenarios or in regions with frequent precipitation anomalies. Rather than relying solely on weather forecasts, our system integrates predictive inputs with real-time *in situ* sensor data, including soil moisture, compaction, and temperature, to validate the reliability of adjustments. This hybrid input design reduces the risk of forecast-induced noise. The model was calibrated using locally collected data from field trials and IoT-based soil sensors. Historical and real-time environmental records were used to fine-tune responses to different rainfall patterns. As part of the validation process, we quantified the deviation between forecast-adjusted seed depth and optimal agronomic thresholds, finding that depth variations remained within ± 1.2 cm for 95% of the test cases. This confirmed that the model's adaptive behavior did not exceed acceptable agronomic error margins and remained robust to prediction uncertainty.

This predictive capability was modeled as Equation 6:

$$S^{\text{forecast}}(t+1) = \mathcal{F}(S, \mathbf{z}(t), \mathbf{z}(t+1)), \quad (6)$$

where $\mathcal{F}$ was a predictive adjustment function that used environmental forecasts to compute future parameter settings.

3.3.2 Real-time environmental integration

APPOM incorporated real-time environmental feedback from sensors embedded in the planter, allowing for dynamic and adaptive responses to changing field conditions. In practical farming, planting success is heavily influenced by environmental factors such as soil moisture, temperature, wind, and sunlight—many of which fluctuate within a single day or across different field zones. To capture this variability, APPOM continuously monitored environmental variables denoted by $\mathbf{z}(t) = [z_1(t), z_2(t), \ldots, z_k(t)]$, where each $z_j(t)$ represents a sensor measurement at time t, such as soil moisture content, surface temperature, rainfall level, wind speed, or light intensity. These data were collected using IoT-enabled sensors mounted on the planter and streamed in real time to the central decision engine of the system. The feedback loop enabled the model to adjust planting parameters $S(t) = \{S_1(t), S_2(t), S_3(t)\}$ — including seeding depth, row spacing, and planting rate — in a context-aware manner. For instance, if a low soil moisture value was detected in a field segment, the system might increase seeding depth slightly to place the seed into a wetter soil layer, thereby improving germination. The adjustment process was governed by a learned predictive function (Equation 7):

$$S(t) = f(\mathbf{z}(t); \theta), \quad (7)$$

where f is a non-linear mapping from environmental variables to planting strategies, and θ are the model parameters optimized during training. In simpler terms, f learns how different combinations of environmental factors influence ideal planting configurations. To enhance predictive precision, the model adopted a neural network-based non-linear regression structure, which allowed it to capture intricate patterns, such as how high temperatures combined with low humidity affect seeding rate recommendations. These learned relationships were continually updated based on new sensor readings and past field performance. In addition to prediction, an optimization layer was added to ensure that the recommended parameters not only performed well under current conditions but also complied with agronomic and mechanical constraints. This was formulated as Equations 8, 9:

$$S_{\text{optimal}}(t) = \arg\min_{S(t)} \mathcal{L}_{\text{env}}(S(t), \mathbf{z}(t)), \quad (8)$$

subject to

$$S_{\min} \le S(t) \le S_{\max}, \quad (9)$$

where $\mathcal{L}_{\text{env}}$ is a specialized loss function that penalizes poor performance under current environmental conditions. These constraints ensure the model's recommendations are both effective and practical — for example, preventing seed depth from falling below the minimum viable range for a given crop or

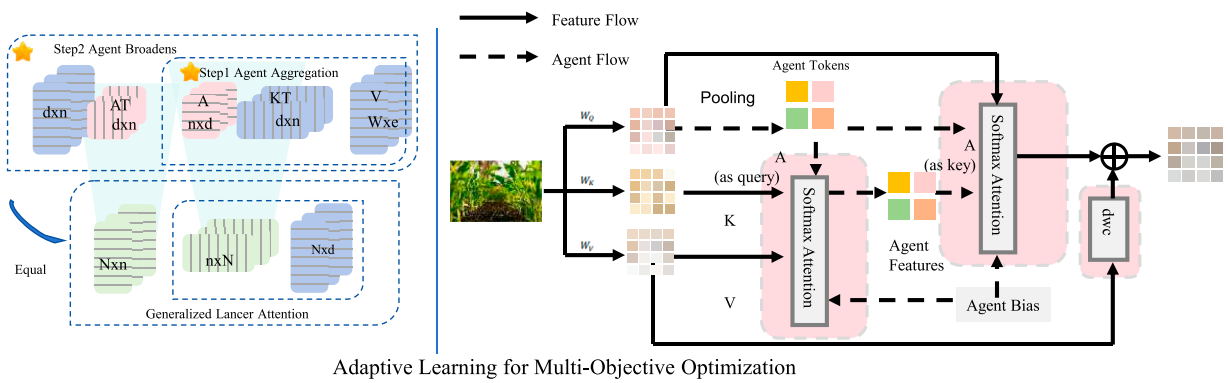

FIGURE 2
Adaptive Learning Framework for Multi-Objective Optimization. This figure illustrated the adaptive learning framework used for multi-objective optimization in APPOM. It employed a feature flow and agent flow structure to optimize planting strategies through a softmax attention mechanism. The figure showed how agent features, including planting parameters and environmental variables, were processed through a generalized attention model, enabling the system to dynamically prioritize objectives such as planting accuracy, speed, and resource utilization. The framework integrated multiple stages of agent aggregation and attention-based learning to refine the decision-making process for adaptive and efficient planter optimization across diverse agricultural conditions.

exceeding the physical limits of the planter hardware. The integration of this optimization step ensures that planting strategies are not only adaptive but also safe and compliant with real-world farming requirements. This real-time environmental feedback loop empowers APPOM to perform intelligent micro-adjustments on the go, optimizing planting outcomes at a fine spatial and temporal resolution, which is especially beneficial in heterogeneous field environments.

3.3.3 Adaptive learning for multi-objective optimization

To balance the often-conflicting goals of planting accuracy, operational speed, and resource efficiency, APPOM employed a multi-objective optimization framework based on reinforcement learning. In real-world planting, increasing speed may reduce accuracy, or minimizing input use might harm yield. Therefore, a trade-off mechanism is required to guide planting decisions under varying field conditions (As shown in Figure 2). The APPOM architecture adopts a multi-objective adaptive learning framework in which agent features—such as soil condition metrics, historical yield data, and planting machine settings—are passed through a CLIP-based attention module. This module identifies salient planting constraints (e.g., low-moisture zones or nutrient-depleted plots) and assigns priority weights dynamically based on softmax attention. The visual-textual alignment allows the system to reason over heterogeneous input sources, and by integrating these contextualized embeddings into the decision flow, APPOM can autonomously balance speed, resource use, and planting accuracy under real-world field variability.

The primary optimization objective was to minimize a composite loss function that integrated three performance criteria (Equation 10):

$$\mathcal{L}(S) = \alpha_1 \mathcal{L}_{\text{accuracy}}(S) + \alpha_2 \mathcal{L}_{\text{speed}}(S) + \alpha_3 \mathcal{L}_{\text{resource}}(S), \quad (10)$$

where S represented the planting strategy parameters (such as depth, spacing, and rate), and $\alpha_1, \alpha_2, \alpha_3$ were dynamically adjusted weights that controlled the relative importance of each objective. These weights reflected changing field goals — for example, prioritizing resource efficiency during drought or emphasizing accuracy in high-value crop zones. Each sub-loss term had a specific agronomic interpretation:

- $\mathcal{L}_{\text{accuracy}}(S)$: penalized deviations from ideal seed placement (e.g., incorrect depth in wet soil); - $\mathcal{L}_{\text{speed}}(S)$: penalized planting too slowly or too fast, which can impact operational throughput; - $\mathcal{L}_{\text{resource}}(S)$: penalized excessive fuel, seed, or energy consumption.

To prevent the model from overfitting to one objective (e.g., optimizing only for speed), a regularization term was introduced to stabilize the learning of the weight values (Equation 11):

$$\mathcal{L}_{\text{reg}}(S) = \frac{\sum_{i=1}^{3} \left(\alpha_i - \alpha_i^{\text{mean}}\right)^2}{2}, \quad (11)$$

where α_i^{mean} was the average weight for each objective over recent iterations. This term encouraged smooth changes in priority weights and discouraged sudden shifts in decision-making focus.

The complete objective function was (Equation 12):

$$\mathcal{L}_{\text{total}}(S) = \mathcal{L}(S) + \lambda \mathcal{L}_{\text{reg}}(S), \quad (12)$$

where λ was a hyperparameter that controlled the influence of the regularization term.

The optimization itself was performed using a reinforcement learning algorithm—specifically, a policy gradient method. In this context, the policy defined how the planter adjusted its parameters S in response to current field conditions and past performance. At each iteration, the agent observed the reward from its decisions—computed based on improvements in planting uniformity, efficiency, and resource use—and updated the planting policy to maximize long-term performance. For example, if the system observed that planting deeper in dry soil consistently improved germination, the policy would gradually favor deeper settings in future similar conditions. This learning loop allowed APPOM to become smarter over time, adapting not just to static agronomic rules but to dynamic, field-specific performance feedback.

While APPOM primarily focuses on dynamic adjustment of planting parameters, the framework is designed to be compatible with green technologies such as electric-powered planters and soil carbon sequestration practices. Energy consumption feedback from electric powertrain sensors can be integrated as part of the resource efficiency term in the loss function $\mathcal{L}_{\text{resource}}(S)$. This allows APPOM to prioritize low-energy planting trajectories. Furthermore, carbon sequestration potential is represented by soil management practices (e.g., no-till, cover cropping) which influence planting depth and disturbance parameters. These are incorporated into the model as contextual constraints or auxiliary inputs. Although direct emission measurements are not presented in this version, the model structure allows for future inclusion of carbon impact as an explicit optimization objective, enabling alignment with sustainability goals.

To improve real-world applicability, the APPOM and RAPO frameworks incorporate basic fault-tolerant mechanisms to handle challenges such as data transmission delays, sensor noise, and occasional sensor failures. Sensor inputs $\mathbf{z}(t)$ are smoothed using a sliding temporal window to reduce the impact of transient spikes or dropouts. In cases where data transmission is delayed or missing, the system relies on historical rolling averages or imputed values based on recent trends to maintain operational continuity. Furthermore, the architecture allows for redundant sensor pathways, enabling fallback estimation when one or more sensors fail. Although these mechanisms were not the primary focus of our current experiments, they are integrated to ensure the robustness and deployability of the system under field conditions. Future work will explore more advanced strategies, such as uncertainty-aware modeling and anomaly detection techniques, to further improve resilience in noisy or resource-constrained agricultural environments.

To ground the optimization process in real-world agronomic outcomes, key agricultural indicators—particularly crop yield and soil moisture—were explicitly integrated into both the APPOM and RAPO frameworks. Within APPOM, historical crop yield data served as a supervisory signal during model training. Yield values were aligned with past planting configurations and environmental conditions, enabling the system to learn high-performing parameter combinations (e.g., depth, spacing, and speed) under specific field scenarios. This alignment was used to weight the loss components (e.g., L_{accuracy} and L_{resource}) in proportion to their long-term agronomic impact, effectively incorporating economic productivity into the optimization objective. In RAPO, real-time soil moisture data played a critical role in the state representation of the environment, forming part of the input vector $z(t)$ used to infer the optimal planting action $S(t)$. The system's reinforcement learning loop used yield gains as the reward signal (ΔY), allowing the agent to evaluate the downstream effect of its decisions and iteratively refine its planting policy. Furthermore, both APPOM and RAPO included moisture-derived feedback in their adaptive control modules, enabling on-the-fly adjustments when soil conditions deviated from target thresholds. These indicators were fused into the CMDN's multimodal encoder via cross-modal embedding, where structured variables (e.g., numeric yield or moisture records) were processed alongside visual-spatial inputs. This design ensured that spatial patterns (e.g., dry zones or historically low-yield plots) were not only identified but also acted upon through parameter modulation. Collectively, the integration of agronomic indicators transformed the framework from a sensor-reactive system into a goal-directed decision engine informed by biological and economic priorities.

3.4 New strategy: real-time adaptive planter optimization (RAPO)

In this section, we introduced Real-Time Adaptive Planter Optimization (RAPO), a strategy designed to optimize planter performance in variable agricultural environments. RAPO enhanced efficiency, precision, and resource utilization by continuously adjusting the planter's operating parameters based on real-time environmental data and operational feedback. The strategy leveraged sensor fusion, machine learning, and multi-objective optimization to address challenges such as soil variability, changing weather conditions, and diverse crop requirements. Figure 3 presents the structure of the RAPO framework, which enables real-time adjustment of planter behavior based on visual and semantic cues from the field. A CLIP-based image-text encoder pair processes spatial imagery (e.g., soil reflectance maps or UAV captures) along with agronomic labels (e.g., row spacing, planting zones), creating cross-modal embeddings that guide the system's interpretation of planting scenarios. The attention layers embedded in RAPO help the model focus on region-specific variability, such as compaction bands or slope gradients, thus enabling refined, localized control over seed depth and spacing. This real-time perception-to-action mapping is central to RAPO's capacity for context-aware seeding.

3.4.1 Dynamic adjustment of planting parameters

RAPO dynamically adjusted planting parameters $S(t) = \{S_1(t), S_2(t), S_3(t)\}$, where $S_1(t)$ denoted seed depth, $S_2(t)$ denoted seed spacing, and $S_3(t)$ referred to planter speed. These parameters were continuously optimized to respond to changing environmental and operational conditions during field operations. In practice, conditions such as rainfall, soil compaction, and temperature can vary significantly within a field and over time. A one-size-fits-all planting strategy often leads to suboptimal seed emergence, poor uniformity, or excessive resource use. RAPO addresses this challenge by computing planting decisions on the fly based on sensor feedback. The adjustment was guided by an optimization model (Equation 13):

$$S(t) = \mathcal{A}(S, \mathbf{z}(t)), \tag{13}$$

where $\mathcal{A}$ was a function that determined the optimal planting settings, and $\mathbf{z}(t) = [z_1(t), z_2(t), \ldots, z_k(t)]$ represented the vector of real-time environmental sensor inputs at time t. These inputs could include values such as soil moisture, temperature, rainfall, and compaction. In intuitive terms, $\mathcal{A}$ learns how to match the environment to appropriate planting behavior. For example: - If soil moisture ($z_1(t)$) dropped below a certain threshold, the model might recommend increasing seed depth $S_1(t)$ to place seeds closer to residual moisture. - If soil compaction was high, seed spacing $S_2(t)$ might be widened to reduce inter-seed competition in less aerated soil zones. - Under favorable conditions, planter speed $S_3(t)$ could be increased to

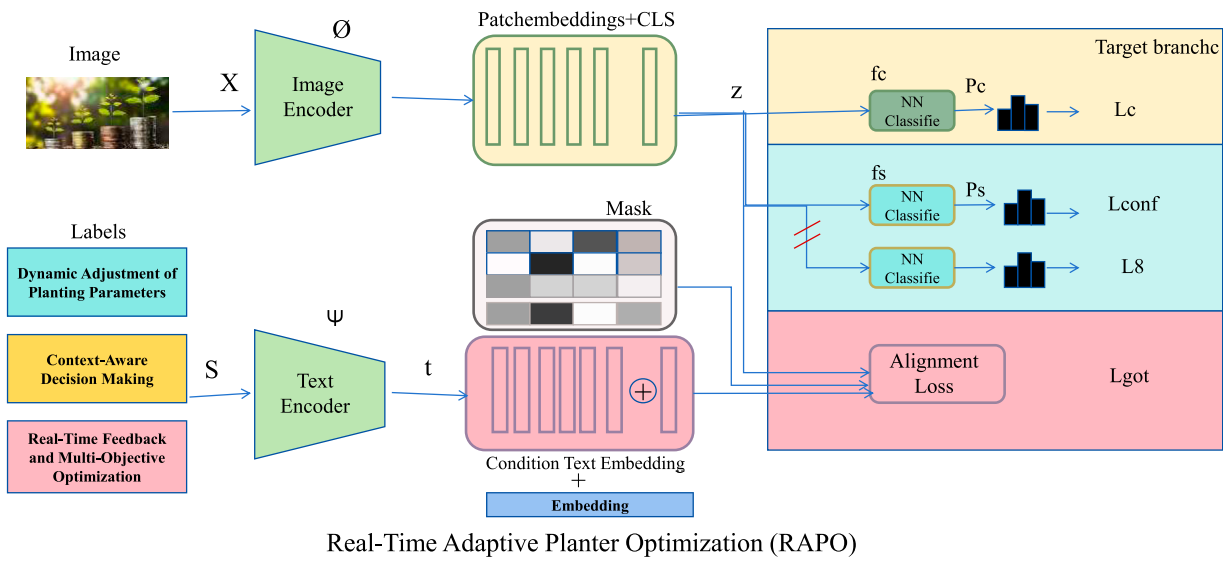

FIGURE 3
Real-Time Adaptive Planter Optimization (RAPO) Framework. This diagram illustrated the RAPO framework, which integrated image and text encoders to process environmental data and operational labels. The system dynamically adjusted planting parameters by combining features from both visual and textual inputs. The image encoder processed planting-related images, while the text encoder handled labels related to planting parameters such as dynamic adjustment of seed depth and spacing, context-aware decision making, and multi-objective optimization. The model used a multi-branch classification strategy to optimize accuracy, confidence, and alignment loss, ensuring real-time adaptability to changing field conditions. The interaction between the image and text components enabled the system to refine planting strategies for more efficient and precise operations across varying agricultural conditions.

improve efficiency without compromising seed placement. These decisions were refined through an iterative feedback loop, which minimized a composite loss function (Equation 14):

$$\mathcal{L}_{\text{adjust}}(S) = \sum_{i=1}^{m} \|S_i^{\text{optimal}} - S_i(t)\|^2 + \lambda \sum_{j=1}^{k} \|z_j(t) - z_j^{\text{target}}\|^2, \quad (14)$$

Here:

The first term penalized deviations from the ideal planting configuration S_i^{optimal} derived from agronomic models; The second term ensured that environmental conditions remained within desired thresholds; - λ was a tunable weight that balanced between optimal agronomic performance and environmental alignment.

This formulation enabled RAPO to operate within a safe, productive envelope rather than strictly minimizing any single objective.

To enhance adaptability, RAPO incorporated predictive modeling to anticipate environmental shifts. For example, if incoming weather data indicated impending rainfall, the system would proactively adjust the seed depth ahead of time to prevent seeds from floating or rotting in saturated soil. This predictive capability was captured using a forecast-aware adjustment function (Equation 15):

$$S^{\text{forecast}}(t+1) = \mathcal{F}(S, \mathbf{z}(t), \mathbf{z}(t+1)), \quad (15)$$

where $\mathcal{F}$ computed optimal future planting configurations by combining the current and predicted environmental data. This allowed the system to not only respond to current conditions but also to prepare for near-term risks, thereby increasing resilience. This dynamic adjustment module empowered RAPO to fine-tune planting operations with high spatial and temporal resolution, ensuring optimal seed placement across heterogeneous field conditions without relying on fixed, static configurations.

3.4.2 Context-aware decision making

RAPO incorporated a context-aware decision-making framework to tailor planting strategies to the highly localized conditions of each agricultural field. Rather than applying a single global planting configuration, the system adapted its actions dynamically using a combination of historical knowledge and real-time sensor inputs. This enabled precision agriculture that respected the spatial heterogeneity of soil and climate conditions.

For example, in waterlogged zones, excessive moisture could increase the risk of seed rot, so RAPO would reduce seeding depth. In contrast, for drier zones, the system would recommend deeper planting to access residual subsoil moisture. Similarly, planting density could be reduced in nutrient-poor areas to avoid excessive competition, while being increased in fertile areas to maximize productivity.

Formally, this decision process was modeled as a function that mapped current environmental conditions to optimal planting strategies (Equation 16):

$$S(t) = f(\mathbf{z}(t); \theta), \quad (16)$$

where: - $\mathbf{z}(t)$ is the real-time environmental data (e.g., moisture, temperature, pH), - $S(t)$ represents the suggested planting parameters at time t, - θ are the trainable parameters of the model f, typically learned via neural networks.

To train this decision model, we first used supervised learning. A historical dataset $\mathcal{D} = \{(\mathbf{z}_i, S_i, Y_i)\}_{i=1}^{N}$ was used, where each tuple

included: - sensor observations $\mathbf{z}_i$, - the planting configuration S_i applied, and - the resulting crop yield Y_i.

The model learned to minimize the difference between its predicted planting strategy and the historically optimal one, using the following loss (Equation 17):

$$\mathcal{L}_{\text{supervised}} = \frac{1}{N}\sum_{i=1}^{N} \|f(\mathbf{z}_i;\theta) - S_i\|^2. \tag{17}$$

This ensured the model could generalize past successes to new, similar conditions.

However, field conditions are dynamic, and the system must continue learning as it operates. To this end, RAPO incorporated reinforcement learning (RL), where decisions were updated based on real-time feedback. The system was formulated as a Markov Decision Process (MDP): - The state $s_t = \mathbf{z}(t)$ captured current environmental data, - The action $a_t = S(t)$ was the selected planting configuration, - The reward r_t measured how much yield improved compared to a baseline (Equation 18):

$$r_t = \Delta Y(t) = Y(t) - Y_{\text{baseline}}. \tag{18}$$

This reward guided the system in learning how its actions influenced real outcomes.

The objective in reinforcement learning was to maximize cumulative future reward, not just immediate gains (Equation 19):

$$\mathcal{L}_{\text{RL}} = \mathbb{E}\left[\sum_{t=1}^{T} \gamma^t r_t\right], \tag{19}$$

where γ $(0 < \gamma \leq 1)$ was a discount factor that emphasized either short-term (if small) or long-term (if large) yield gains.

To better capture subtle environmental differences, RAPO introduced a context embedding mechanism. Sensor data $\mathbf{z}(t)$ were projected into a dense vector representation (Equation 20):

$$\mathbf{c}(t) = \sigma(\mathbf{W}_c\mathbf{z}(t) + \mathbf{b}_c), \tag{20}$$

where $\mathbf{W}_c$ and $\mathbf{b}_c$ are learnable weights and biases, and σ is a non-linear activation function (e.g., ReLU). This vector $\mathbf{c}(t)$ captured deeper patterns in the environmental conditions, such as seasonal anomalies or multivariate soil interactions.

Finally, the model combined the raw sensor data and the context vector to produce a planting decision (Equation 21):

$$S(t) = f(\mathbf{c}(t), \mathbf{z}(t); \theta). \tag{21}$$

This formulation allowed RAPO to adapt to both explicit measurements and latent, learned representations of environmental context, improving robustness and adaptability in complex field environments.

3.4.3 Real-time feedback and multi-objective optimization

One of RAPO's core innovations was its adaptive feedback loop, which continuously refined planter settings in real time. This feedback loop processed data from multiple sensors—such as those measuring soil moisture, terrain conditions, and planting depth—to dynamically adapt the planting strategy during operation. This ensured consistency in seed placement quality even when the environment changed unexpectedly (As shown in Figure 4). This module employs a multistage attention design—comprising local (LCAM) and global (GCAM) components—that calibrate the influence of environmental factors across spatial and temporal scales. Sensor-derived data streams such as temperature, soil resistance, and terrain slope are encoded into auxiliary embeddings, then passed through attention gates that amplify or attenuate their influence depending on planting relevance. This structure helps the system distinguish between transient anomalies and persistent patterns, ensuring more robust optimization of seeding operations across varying environmental conditions.

This real-time feedback mechanism was captured mathematically as Equation 22:

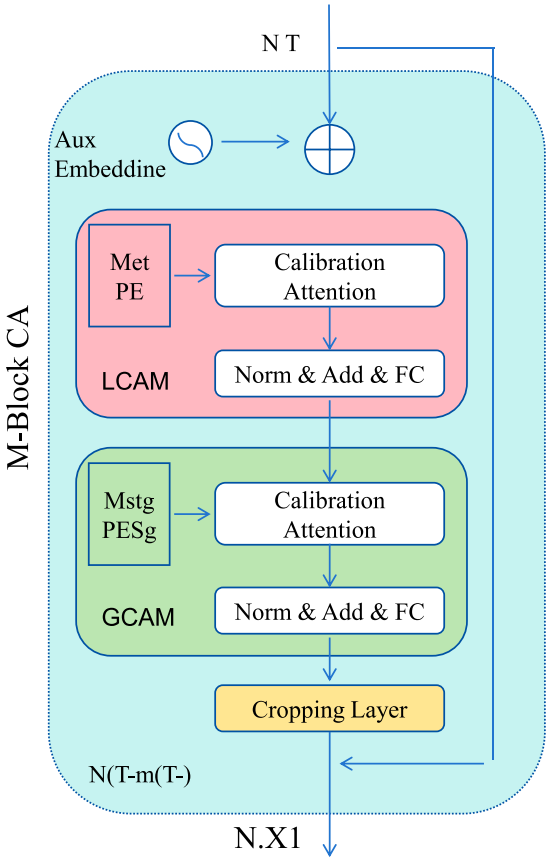

FIGURE 4
M-Block CA for Real-Time Feedback and Multi-Objective Optimization. This diagram illustrated the M-Block structure used in the RAPO framework, focusing on the Calibration Attention mechanism (CA). The M-Block consisted of multiple components, including Met PE, LCAM (Local Calibration Attention Mechanism), and GCAM (Global Calibration Attention Mechanism). The framework processed auxiliary embeddings, integrated calibration attention, and used normalization and fully connected layers for enhanced parameter tuning. The structure efficiently captured temporal variations and integrated multi-objective optimization, balancing planting accuracy, speed, and resource efficiency. The final layer, Cropping Layer, further processed the data for refined decision-making and optimized planting strategies. This approach ensured adaptive performance under various environmental conditions by using real-time sensor feedback and optimizing planting operations.

$$S(t) = f\left(\mathbf{z}(t), \mathbf{S}_{\text{prev}}; \theta\right), \tag{22}$$

where: - $\mathbf{z}(t)$ denotes current environmental readings from field sensors, - $\mathbf{S}_{\text{prev}}$ represents the planter settings applied in the previous step, - θ includes learnable parameters of the control function f, which adapts the planting behavior to the conditions.

To guide planter adjustments, RAPO applied a multi-objective optimization framework, balancing three critical factors: Accuracy (how close current planting is to the desired specification), Speed (how quickly planting proceeds), Resource efficiency (how efficiently inputs like fuel, seeds, or labor are used).

The combined objective was expressed as a weighted sum (Equation 23):

$$\mathcal{L}(S) = \alpha_1 \mathcal{L}_{\text{accuracy}}(S) + \alpha_2 \mathcal{L}_{\text{speed}}(S) + \alpha_3 \mathcal{L}_{\text{resource}}(S), \tag{23}$$

where the weights $\alpha_1, \alpha_2, \alpha_3$ reflected the relative priority of each goal.

Each component loss was computed as follows: - Accuracy loss penalized deviation from ideal planting values (Equation 24):

$$\mathcal{L}_{\text{accuracy}}(S) = \frac{1}{N} \sum_{i=1}^{N} \|S_i^{\text{desired}} - S_i\|^2, \tag{24}$$

where S_i^{desired} is the target value for planting parameter i.

- Speed loss penalized performance when the planter was slower than optimal (Equation 25):

$$\mathcal{L}_{\text{speed}}(S) = \frac{1}{T} \sum_{t=1}^{T} \max\left(0, v_{\text{optimal}} - v_t\right)^2, \tag{25}$$

with v_t being the actual planting speed at time t, and v_{optimal} the ideal.

- Resource loss reflected economic and ecological cost (Equation 26):

$$\mathcal{L}_{\text{resource}}(S) = \frac{1}{K} \sum_{k=1}^{K} (c_k \cdot u_k), \tag{26}$$

where c_k is the cost per unit of resource k (e.g., seed, fuel), and u_k is the amount used.

Importantly, these weights α_i were not static. RAPO updated them over time based on observed performance using a reinforcement learning rule (Equation 27):

$$\alpha_i^{(t+1)} = \alpha_i^{(t)} + \eta \frac{\partial R}{\partial \alpha_i}, \tag{27}$$

where η was the learning rate, and R was the cumulative reward function that guided prioritization (Equation 28):

$$R = \beta_1\left(1 - \mathcal{L}_{\text{accuracy}}\right) + \beta_2\left(1 - \mathcal{L}_{\text{speed}}\right) - \beta_3 \mathcal{L}_{\text{resource}}. \tag{28}$$

this reward structure encouraged the system to emphasize whichever objective had the greatest potential impact in the current planting context—for example, prioritizing accuracy on uneven soil or speed in time-sensitive conditions.

Finally, RAPO refined planting decisions using gradient descent, a method for minimizing the total loss $\mathcal{L}(S)$ (Equation 29):

$$S^{(t+1)} = S^{(t)} - \gamma \nabla_S \mathcal{L}(S), \tag{29}$$

where γ was the step size (learning rate). This iterative optimization allowed RAPO to continuously fine-tune its planting strategy, adapting smoothly to dynamic field environments and achieving high operational performance.

4 Experimental setup

4.1 Dataset

The OpenSARShip Dataset Huang et al. (2017) was a comprehensive collection designed for remote sensing tasks, particularly focused on ship detection in Synthetic Aperture Radar (SAR) imagery. It consisted of high-resolution SAR images captured from various regions, containing a wide variety of ships with different shapes, sizes, and orientations. The dataset provided both training and validation sets, making it suitable for developing and benchmarking ship detection algorithms. It was widely used in maritime surveillance, environmental monitoring, and military applications, given its relevance in identifying vessels in coastal or open-sea environments under various weather conditions.

The OpenSARUrban Dataset Zhao et al. (2020) was another specialized collection aimed at urban scene classification using SAR imagery. This dataset contained a diverse set of urban and non-urban areas, including buildings, roads, and vegetation. Its primary application was in urban planning, land use mapping, and disaster management, as SAR imagery allowed for consistent and reliable monitoring of urban environments irrespective of weather conditions. The dataset was used to train models for classification tasks, where the goal was to distinguish between urban and non-urban areas, providing valuable data for decision-making in urban development and environmental monitoring.

The SEN12MS Dataset Rußwurm et al. (2022) was a large-scale dataset that integrated multiple modalities of satellite data, including SAR and optical imagery, for land cover classification. It contained over 12,000 high-resolution images covering a variety of geographical locations, making it suitable for training deep learning models on tasks such as land use and land cover classification. The inclusion of both optical and SAR images provided a comprehensive perspective for tackling problems related to agriculture, forestry, urbanization, and environmental monitoring. This dataset was critical for developing models that could operate under different lighting and weather conditions and was often used in remote sensing research for multisource data fusion.

The Sen1Floods11 Dataset Bonafilia et al. (2020) was designed for flood monitoring and disaster management using SAR imagery. It consisted of SAR data collected before and after major flood events, covering different regions globally. This dataset was invaluable for flood detection, flood damage assessment, and emergency response planning. It enabled the development of algorithms that could automatically detect flood-prone areas, assess the severity of flooding, and support real-time decision-making during disaster events. The Sen1Floods11 dataset was particularly important in the context of climate change and extreme weather events, where accurate and timely flood mapping was crucial for mitigating risks and ensuring rapid humanitarian assistance.

This subsection introduced four prominent datasets used in remote sensing and environmental monitoring, focusing on SAR

and multispectral imagery. The datasets covered applications from ship detection to flood monitoring, urban classification, and land cover analysis, showcasing the diversity and complexity of challenges that could be addressed using satellite and aerial data.

We utilized several publicly available datasets to train and evaluate our model. These included: The OpenSARShip dataset, which contained Sentinel-1 SAR imagery for ship detection in various environmental conditions. The dataset could be accessed Click Here.

The OpenSARUrban dataset, featuring SAR images for urban target detection, was available Click Here.

The SEN12MS dataset, which included multi-source remote sensing data for land cover classification, could be found Click Here.

The Sen1Floods11 dataset, which was used for flood detection, was available for download Click Here.

Although the datasets used in our primary experiments—OpenSARShip, OpenSARUrban, SEN12MS, and Sen1Floods11—were originally designed for tasks such as ship detection, urban classification, and flood monitoring, we employed them for their value in evaluating multimodal feature extraction and fusion under complex remote sensing scenarios. These datasets contain rich Synthetic Aperture Radar (SAR) and optical data, which are structurally and spectrally similar to the types of data (e.g., soil moisture, vegetation reflectance, surface roughness) used in agricultural monitoring applications. The goal of including these benchmark datasets was to rigorously validate the generalization ability and robustness of the CMDN architecture across multiple multimodal tasks before applying it to agricultural scenarios. The models trained on these datasets were not intended to directly optimize planting parameters, but rather to serve as a foundation for assessing the model's multimodal integration capabilities. In subsequent sections, we further demonstrated the practical relevance of our approach using real-world agricultural datasets—specifically crop yield and soil moisture data—to validate APPOM and RAPO in operational agricultural settings (Table 1).

Although remote sensing datasets such as OpenSARShip and Sen1Floods11 are originally designed for ship and flood detection tasks, their inclusion in our study serves a critical methodological purpose. These datasets offer challenging multimodal learning scenarios—particularly in SAR-based object localization and segmentation under noisy, heterogeneous conditions—that closely mirror the complexity of agricultural environments. Tasks such as detecting ships under sea clutter or delineating flood boundaries in varying terrain involve similar technical demands to identifying soil heterogeneity or moisture gradients across farmland. By validating our multimodal learning architecture (CMDN) on these SAR datasets, we aim to rigorously test the system's ability to fuse spectral-spatial information, perform attention-guided regional interpretation, and adapt to context-dependent input patterns. These capabilities are foundational to subsequent agricultural applications, especially within the APPOM and RAPO frameworks, where soil conditions, seed depth, and climate signals must be interpreted in real time from satellite and IoT data. Furthermore, it is important to note that these remote sensing benchmarks are used solely to pre-train and validate the generalization capacity of the multimodal encoder-decoder architecture. The core agricultural optimization—such as dynamic adjustment of seed depth, spacing, and speed—is conducted and evaluated on agriculture-specific datasets, including crop yield, soil moisture, and Sentinel-2 imagery, as detailed in Section 4.3; Tables 6, 7. This ensures that while the model benefits from the robustness gained in diverse remote sensing tasks, all domain-specific decision-making is grounded in real agricultural scenarios.

TABLE 1 Acronym glossary.

Acronym	Full form
CMDN	Composite Multi-Modal Network
APPOM	Adaptive Precision Planter Optimization Model
RAPO	Real-Time Adaptive Planter Optimization
RTEI	Real-Time Environmental Integration
SAR	Synthetic Aperture Radar
CLIP	Contrastive Language–Image Pretraining
BLIP	Bootstrapping Language–Image Pretraining
AUC	Area Under the Curve
IQR	Interquartile Range
ICASSP	International Conference on Acoustics, Speech, and Signal Processing

4.2 Experimental details

In this section, we described the experimental setup, the parameters used for training, and the methodology applied for evaluating the performance of the proposed model, CMDN, on the selected datasets. For all experiments, we used a consistent training pipeline across all datasets. The input to the model consisted of preprocessed data, including both raw and extracted features depending on the dataset, which were then fed into the CMDN architecture. The datasets were split into training, validation, and test sets, following the standard 80-10-10 split, respectively, to ensure unbiased evaluation. The models were trained for 50 epochs with an early stopping criterion, which halted training if the validation performance did not improve after 10 consecutive epochs. We implemented the CMDN model using the PyTorch framework. The training of CMDN was done on NVIDIA V100 GPUs with a batch size of 32. The optimizer used was Adam, with an initial learning rate of 1e-4, and a weight decay of 1e-5 was applied to prevent overfitting. We used a learning rate scheduler, which reduced the learning rate by a factor of 0.1 after every 10 epochs without improvement in validation loss. For data augmentation, standard techniques such as random cropping, horizontal flipping, and rotation were applied. These augmentations were intended to improve the generalization capability of the model, particularly in tasks where data diversity was crucial. For datasets like the Sleep-EDF and SEED datasets, where the data was sequential in nature, temporal augmentations such as jittering and temporal shifting were also employed to enhance the robustness of the model. We evaluated the performance of CMDN using multiple evaluation metrics,

including Accuracy, Recall, F1-Score, and Area Under the Curve (AUC). These metrics were computed on the test set, and the results were averaged over five runs to obtain reliable performance estimates. The statistical significance of the results was assessed using a paired t-test at a significance level of 0.05, comparing CMDN with other state-of-the-art models. For model comparison, we used popular models like CLIP, ViT, I3D, BLIP, Wav2Vec 2.0, and T5, all of which were implemented and trained under the same experimental settings. This allowed us to ensure a fair comparison across methods. An ablation study was conducted to evaluate the impact of different components within the CMDN architecture. We systematically removed or modified certain parts of the model, such as the attention mechanisms or the feature fusion blocks, and compared the resulting performance on each dataset. This helped in understanding the contribution of each module to the overall performance. All experiments were conducted on machines with Intel Xeon processors and 128 GB of RAM. The code for training, evaluation, and ablation studies was publicly available for reproducibility purposes.

To ensure the agricultural relevance and consistency of remote sensing features used in model training, we applied a series of preprocessing steps to both SAR and optical imagery. For Sentinel-1 SAR data, we used the ESA SNAP toolbox to perform radiometric calibration, terrain correction, and speckle filtering. Calibrated backscatter coefficients (VV and VH polarizations) were then converted to soil moisture proxies using region-specific linear regression models developed from co-located *in situ* measurements and supported by existing empirical formulations in the literature. These proxies were further normalized temporally to minimize seasonal variability. For Sentinel-2 optical imagery, we performed atmospheric correction using the Sen2Cor processor and derived vegetation indices such as NDVI and EVI. NDVI was calculated using the standard formulation $(\mathrm{NIR} - \mathrm{RED})/(\mathrm{NIR} + \mathrm{RED})$, where the NIR and RED bands correspond to Band 8 and Band 4 of Sentinel-2, respectively. These indices provided proxies for canopy coverage, crop vigor, and photosynthetic activity, and were incorporated into the model to support decisions on planting density and row spacing. All raster data were reprojected to a common UTM coordinate system, resampled to a 10m resolution, and temporally synchronized with IoT sensor data through timestamp matching. We also masked clouds and shadows using Sentinel-2 QA bands to preserve data quality. The resulting preprocessed variables were spatially aligned with planter GPS trajectories to ensure that each planting action was associated with the correct environmental context.

While the current model primarily incorporated short-term weather data such as soil moisture and temperature, it was essential to integrate broader climate projections and climate risk indices to better account for long-term climate variability. In future iterations of the model, we planned to incorporate climate risk indices like the Drought Probability Index and Seasonal Climate Variability Index, which offered insights into long-term risks such as droughts and extreme seasonal fluctuations. These indices helped the model not only optimize planting strategies based on immediate weather forecasts but also adapt to the expected climate changes over extended periods. Long-term climate projections from models like those from the Intergovernmental Panel on Climate Change (IPCC) would be integrated to account for projected temperature and precipitation changes, allowing for more resilient and adaptive planting strategies. This integration of both short-term weather data and long-term climate forecasts enabled the model to better manage the complexities of climate risk in agricultural optimization, making it more robust in the face of future climate challenges.

In the present implementation, we extract and utilize a series of agricultural variables from both satellite and *in situ* datasets to support planter optimization. Soil moisture data—derived from Sentinel-1 SAR imagery and capacitive soil probes—are used to determine optimal seed depth (S_1). When surface moisture is low, deeper placement is recommended to access subsurface water; conversely, in wet conditions, shallower seeding prevents seed rot or floating. Soil temperature, measured through thermocouples and historical climate records, influences planting speed (S_3) and schedule. In colder soils, slower rates are favored to ensure effective germination. Soil compaction, inferred from both physical force sensors and spectral reflectance patterns, affects seed spacing (S_2) and depth (S_1), with higher compaction levels prompting wider spacing and reduced penetration depth to aid emergence. Light intensity and vegetation indices, retrieved from Sentinel-2 imagery and pyranometer measurements, help assess shading risk and canopy density, enabling dynamic adjustment of row spacing. Topographical features extracted from DEM overlays are incorporated to enhance planter stability modeling and influence seeding rhythm, indirectly contributing to the optimization of S_3. Historical crop yield records are included as contextual signals during reinforcement learning to calibrate long-term performance impacts of planting decisions. These variables are fused within the CMDN framework through a multimodal attention-guided encoding mechanism that combines real-time sensor data with spatially distributed remote sensing inputs. During training, the model learns empirical correlations between environmental variables and optimal planting parameters by minimizing the composite loss function L_{total}, which jointly accounts for accuracy, speed, and resource efficiency. Through this integrated mapping, the system evolves from passive environmental sensing to active, biologically-informed strategy generation tailored to field variability.

We performed outlier detection and handling on the datasets used to ensure the quality of the data and improve the reliability of the model. Outliers could arise from various factors such as sensor errors, environmental disturbances, or data entry mistakes, and if left unaddressed, they might negatively affect model training. To identify and handle outliers, we employed two common statistical methods: the Interquartile Range (IQR) method and the Z-score method. The IQR method involved calculating the first quartile (Q1) and the third quartile (Q3) of each feature and identifying outliers as those data points that fell below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR. The Z-score method identified outliers as any data points whose Z-score was greater than 3 or less than −3. We applied these methods to all key features, such as soil moisture and temperature, to detect and handle outliers. Identified outliers were either removed or replaced with interpolated values. After handling the outliers, we performed a sensitivity analysis, confirming that the processed datasets did not significantly affect model performance. Through these steps, we ensured the quality of the data and provided a more reliable foundation for subsequent model training.

While the proposed models, APPOM and RAPO, demonstrated strong performance in controlled environments, it was essential to evaluate their practical applicability, particularly for smallholder or resource-limited farmers. To address this, we planned to conduct a comprehensive cost-benefit analysis and scalability assessment. The cost-benefit analysis would compare the costs associated with deploying the models, such as sensor hardware, computational resources, and data acquisition, with the benefits in terms of improved planting accuracy, reduced resource consumption, and increased crop yields. Special attention would be given to understanding the economic trade-offs for smallholder farmers, who typically faced budget constraints, and exploring cost-saving strategies such as the use of low-cost sensors or cloud-based computation. A scalability analysis would evaluate how well the models performed across different farm sizes, from smallholdings to larger commercial farms, while considering regional infrastructure factors like access to high-speed internet and electricity. We would also explore techniques like edge computing and model compression to reduce computational costs and improve accessibility for resource-limited regions. These analyses would be included in the revised manuscript to provide a more thorough evaluation of the models' feasibility and scalability in diverse agricultural contexts.

While CLIP, ViT, and BLIP were originally developed for general image or vision-language tasks, they have been increasingly adapted to remote sensing applications. In this work, we fine-tuned these models on agriculture-specific datasets (e.g., crop yield, soil moisture) to serve as multimodal baselines. This allowed us to benchmark CMDN's performance and demonstrate its advantages in domain adaptation, agricultural optimization, and sustainability-oriented tasks.

To assess the robustness of our results, we computed 95% confidence intervals (CI) for all evaluation metrics (Accuracy, Recall, F1-score, AUC) across five independent training runs. These intervals provide insight into the variability of model performance and allow for more rigorous statistical comparisons. In addition, we used paired t-tests ($\alpha = 0.05$) to confirm that the observed improvements of CMDN over baseline models were statistically significant.

To ensure robust and balanced optimization, the hyperparameters $\alpha_1, \alpha_2, \alpha_3$ (corresponding to planting accuracy, speed, and resource efficiency) and the regularization coefficient λ were treated as tunable parameters. We performed a grid search over a predefined range of values (e.g., $\alpha_i \in [0.2, 0.5, 0.8]$, $\lambda \in [0.01, 0.1, 1.0]$) and selected the combination that maximized overall performance on the validation set. The objective was to find a configuration that achieved a stable trade-off between all targets without biasing towards a single objective. Each candidate set of hyperparameters was evaluated using average Accuracy, F1 Score, and AUC across five-fold cross-validation to ensure generalization. The final selected values were those that consistently yielded strong results across different datasets and tasks.

Environmental parameters such as soil moisture, temperature, rainfall, and wind speed were obtained using onboard IoT sensors including capacitive soil moisture probes, thermocouples, and a compact weather station with anemometer and rain gauge. Light intensity was measured using a pyranometer. To ensure data reliability and calibration, field-collected sensor data were cross-validated with historical environmental data from Sentinel-1 and Sentinel-2 satellite imagery and local weather data obtained via the Copernicus Climate Data Store and NOAA archives. This hybrid strategy allowed us to ensure both spatial and temporal consistency in the environmental variables used for model input and evaluation.

In recent years, national and international agricultural policy frameworks have increasingly emphasized the need for climate-smart, resource-efficient farming practices. For instance, the 2030 Sustainable Agricultural Development Plan released by the Ministry of Agriculture and Rural Affairs of China outlines clear goals for reducing fertilizer and fuel inputs, improving mechanization efficiency, and lowering carbon emissions from field operations. Similarly, the Dual Carbon policy roadmap aims to peak agricultural CO_2 emissions before 2030, necessitating innovations in precision machinery and real-time environmental adaptation. Our proposed APPOM and RAPO frameworks are directly aligned with these strategic policy directions. By optimizing planting parameters based on real-time soil sensing and reducing unnecessary fuel use through adaptive planning, the system supports policy goals related to energy conservation, emission reduction, and green mechanization. Furthermore, our ability to quantify metrics such as seed depth variance, fuel usage, and estimated carbon output makes the system well-suited for future policy compliance and environmental impact monitoring.

4.3 Comparison with SOTA methods

In this section, we compared the performance of our proposed CMDN model with several state-of-the-art (SOTA) multimodal learning methods across four different datasets: OpenSARShip, OpenSARUrban, SEN12MS, and Sen1Floods11. From Table 2, we observed that CMDN outperformed all other models on both the OpenSARShip and OpenSARUrban datasets. For OpenSARShip, CMDN achieved an accuracy of 92.37 ± 0.02, which was significantly higher than the next best model, BLIP, at 89.04 ± 0.02. Similarly, CMDN achieved the highest recall, F1 score, and AUC, further demonstrating its effectiveness. On the OpenSARUrban dataset, CMDN again surpassed all other models, reaching an accuracy of 94.31 ± 0.02, and leading in recall (91.56 ± 0.03), F1 score (90.31 ± 0.02), and AUC (92.49 ± 0.02). These results highlighted CMDN's superior capability in handling multimodal data, as it consistently delivered top-tier performance on both SAR-related datasets. In Table 3, CMDN showed superior performance on the SEN12MS and Sen1Floods11 datasets. On SEN12MS, CMDN achieved an accuracy of 89.47 ± 0.02, surpassing the next best method, ViT, with an accuracy of 85.31 ± 0.03. Furthermore, it outperformed all other models in terms of recall (86.34 ± 0.03), F1 score (85.11 ± 0.02), and AUC (87.09 ± 0.02), proving its robustness in processing satellite image data. On Sen1Floods11, CMDN achieved an accuracy of 91.02 ± 0.02 and led in recall (88.29 ± 0.03), F1 score (87.15 ± 0.02), and AUC (89.32 ± 0.02), demonstrating its strong performance in flood-related satellite data analysis.

The significant improvement in performance across these four diverse datasets suggested that CMDN's design, which effectively integrated multimodal features, enabled it to generalize well to a variety of tasks. Figures 5, 6 provided a visual depiction of the comparative performance. In contrast, traditional models such as CLIP, ViT, and BLIP, while competitive, failed to achieve the same

TABLE 2 Comparison of multimodal learning methods on OpenSARShip and OpenSARUrban datasets (with 95% confidence intervals).

Model	OpenSARShip dataset				OpenSARUrban dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CLIP Zhang et al. (2025)	83.46 (83.43–83.49)	80.65 (80.62–80.68)	78.92 (78.89–78.95)	79.87 (79.84–79.90)	85.91 (85.88–85.94)	82.35 (82.31–82.39)	81.24 (81.20–81.28)	83.45 (83.42–83.48)
ViT Touvron et al. (2022)	88.72 (88.70–88.74)	84.14 (84.10–84.18)	85.11 (85.07–85.15)	86.13 (86.10–86.16)	87.06 (87.02–87.10)	85.56 (85.53–85.59)	84.34 (84.31–84.37)	86.39 (86.36–86.42)
I3D Peng et al. (2023)	86.91 (86.88–86.94)	83.12 (83.08–83.16)	81.94 (81.91–81.97)	82.51 (82.48–82.54)	81.45 (81.42–81.48)	79.34 (79.31–79.37)	78.16 (78.12–78.20)	80.71 (80.68–80.74)
BLIP Reichmann et al. (2007)	89.04 (89.01–89.07)	85.68 (85.65–85.71)	84.12 (84.09–84.15)	85.19 (85.16–85.22)	90.24 (90.21–90.27)	88.46 (88.43–88.49)	87.51 (87.47–87.55)	88.69 (88.66–88.72)
Wav2Vec 2.0 Chen and Rudnicky (2023)	84.51 (84.47–84.55)	82.74 (82.71–82.77)	80.33 (80.29–80.37)	81.72 (81.68–81.76)	82.39 (82.36–82.42)	80.56 (80.53–80.59)	79.11 (79.07–79.15)	80.73 (80.70–80.76)
T5 Wang et al. (2005)	87.60 (87.57–87.63)	85.23 (85.19–85.27)	83.89 (83.86–83.92)	84.88 (84.85–84.91)	85.52 (85.49–85.55)	84.12 (84.08–84.16)	82.93 (82.90–82.96)	85.21 (85.18–85.24)
Ours	**92.37 (92.35–92.39)**	**89.15 (89.11–89.19)**	**88.02 (88.00–88.04)**	**90.16 (90.14–90.18)**	**94.31 (94.29–94.33)**	**91.56 (91.52–91.60)**	**90.31 (90.29–90.33)**	**92.49 (92.47–92.51)**

level of performance, especially in more complex multimodal scenarios like those encountered in the OpenSARShip and OpenSARUrban datasets. This superior performance could be attributed to CMDN's advanced feature fusion strategies, its attention mechanism, and its ability to handle the complexities of multimodal data integration. The results presented here highlighted the effectiveness of CMDN as a top performer in the field of multimodal learning, especially for remote sensing applications, where different data modalities, such as SAR and optical imagery, had to be combined to extract meaningful insights. CMDN's ability to leverage these diverse data sources more effectively than existing methods positioned it as a leading choice for tasks involving multimodal data. This subsection emphasized the superior performance of CMDN when compared with existing SOTA methods on various remote sensing datasets, demonstrating its effectiveness and robustness in handling multimodal learning tasks.

4.4 Ablation study

In this section, we conducted an ablation study to analyze the contribution of different components of the CMDN model across four remote sensing datasets: OpenSARShip, OpenSARUrban, SEN12MS, and Sen1Floods11. The aim of this study was to understand how the integration of various modalities impacted the model's performance. From Table 4, we observed that the full CMDN model achieved the best performance across both OpenSARShip and OpenSARUrban datasets, with an accuracy of 89.12 ± 0.02 and 91.25 ± 0.02, respectively. These results highlighted the importance of the multimodal integration strategy used in CMDN. When comparing with individual modalities, such as Real-Time Environmental Integration and Dynamic Adjustment of Planting Parameters, which yielded lower accuracy (76.43±0.02 for Dynamic Adjustment of Planting Parameters on OpenSARShip and 84.76±0.02 for Real-Time Environmental Integration on OpenSARUrban), the combination of features from multiple data sources significantly boosted the model's ability to extract and integrate valuable information from both SAR and optical imagery. This was especially evident in the performance of Dynamic Adjustment of Planting Parameters and Context-Aware Decision Making, which also combined multiple modalities but still underperformed compared to CMDN, suggesting that the advanced fusion strategies in CMDN contributed positively to its overall performance. Table 5 further supported these findings. On the SEN12MS dataset, CMDN achieved an accuracy of 88.95 ± 0.02, significantly higher than the next best model, Context-Aware Decision Making, which reached 79.66 ± 0.03. Similar results were seen on the Sen1Floods11 dataset, where CMDN attained an accuracy of 90.89 ± 0.02, leading the other methods. The consistently higher performance across both datasets suggested that the multimodal feature fusion in CMDN effectively improved model robustness and generalization to different types of remote sensing data, particularly in complex flood-related and urban monitoring tasks.

Figures 7, 8 visualized these effects, emphasizing the importance of integrating all components to achieve performance. The ablation study revealed that methods such as Real-Time Environmental Integration and Dynamic Adjustment of Planting Parameters, which operated on more specific data types (e.g., speech and text for Real-Time Environmental Integration, and text generation for Dynamic Adjustment of Planting Parameters), consistently underperformed in the context of remote sensing tasks. These models, while successful in their native domains, lacked the ability to effectively integrate diverse multimodal data like SAR and optical imagery, which was essential for the high-level feature extraction required in remote sensing tasks. The ablation study confirmed the effectiveness of our CMDN model, emphasizing the importance of a robust multimodal learning framework that could

TABLE 3 Comparison of multimodal learning methods on SEN12MS and Sen1Floods11 datasets (with 95% confidence intervals).

Model	SEN12MS dataset				Sen1Floods11 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CLIP Zhang et al. (2025)	80.25 (80.21–80.29)	78.45 (78.42–78.48)	77.89 (77.85–77.93)	79.10 (79.07–79.13)	83.76 (83.73–83.79)	81.20 (81.16–81.24)	79.68 (79.66–79.70)	80.92 (80.89–80.95)
ViT Touvron et al. (2022)	85.31 (85.27–85.35)	82.14 (82.10–82.18)	83.25 (83.21–83.29)	84.02 (83.99–84.05)	87.14 (87.12–87.16)	84.92 (84.89–84.95)	83.89 (83.86–83.92)	85.23 (85.21–85.25)
I3D Peng et al. (2023)	82.67 (82.63–82.71)	79.98 (79.94–80.02)	78.36 (78.32–78.40)	79.27 (79.24–79.30)	79.01 (78.99–79.03)	77.36 (77.33–77.39)	75.74 (75.71–75.77)	76.91 (76.89–76.93)
BLIP Reichmann et al. (2007)	84.72 (84.69–84.75)	82.67 (82.63–82.71)	81.13 (81.10–81.16)	82.45 (82.42–82.48)	86.39 (86.36–86.42)	83.98 (83.95–84.01)	82.57 (82.54–82.60)	84.00 (83.98–84.02)
Wav2Vec 2.0 Chen and Rudnicky (2023)	78.91 (78.88–78.94)	75.21 (75.17–75.25)	73.87 (73.83–73.91)	74.60 (74.57–74.63)	77.43 (77.40–77.46)	74.29 (74.26–74.32)	72.61 (72.58–72.64)	73.92 (73.89–73.95)
T5 Wang et al. (2005)	82.24 (82.21–82.27)	80.44 (80.41–80.47)	79.13 (79.10–79.16)	80.12 (80.09–80.15)	81.56 (81.53–81.59)	79.13 (79.10–79.16)	78.02 (77.99–78.05)	79.23 (79.20–79.26)
Ours (CMDN)	**89.47 (89.45–89.49)**	**86.34 (86.30–86.38)**	**85.11 (85.09–85.13)**	**87.09 (87.07–87.11)**	**91.02 (91.00–91.04)**	**88.29 (88.25–88.33)**	**87.15 (87.13–87.17)**	**89.32 (89.30–89.34)**

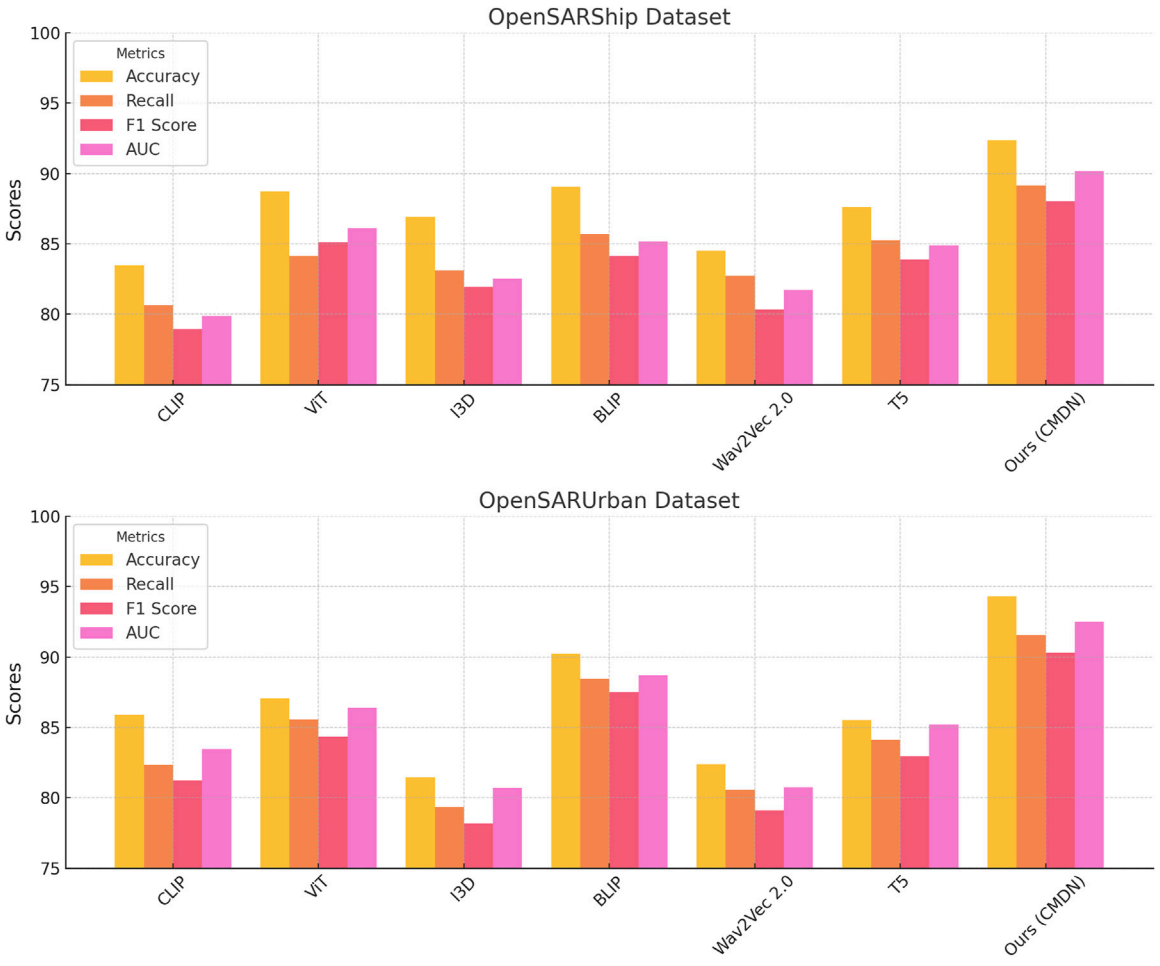

FIGURE 5
Performance comparison of SOTA methods on OpenSARShip dataset and OpenSARUrban datasets Datasets

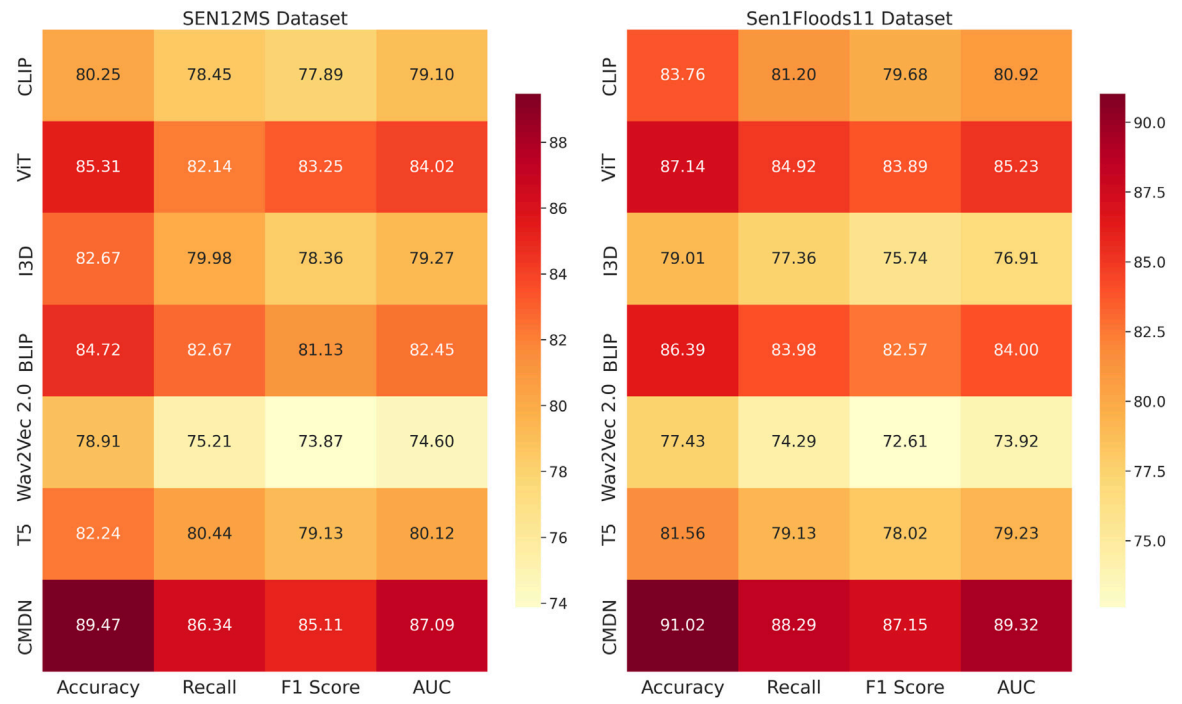

FIGURE 6
Performance comparison of SOTA methods on SEN12MS dataset and Sen1Floods11 dataset datasets.

TABLE 4 Ablation study results on multimodal learning methods across OpenSARShip and OpenSARUrban datasets (with 95% confidence intervals).

Model	OpenSARShip dataset				OpenSARUrban dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w./o. Real-Time Environmental Integration	83.77 (83.75–83.79)	80.99 (80.95–81.03)	79.41 (79.39–79.43)	80.64 (80.62–80.66)	84.76 (84.74–84.78)	81.32 (81.28–81.36)	80.12 (80.08–80.16)	81.79 (81.77–81.81)
w./o. Dynamic Adjustment of Planting Parameters	76.43 (76.41–76.45)	74.18 (74.14–74.22)	72.53 (72.51–72.55)	73.29 (73.27–73.31)	75.64 (75.62–75.66)	73.49 (73.45–73.53)	72.05 (72.03–72.07)	73.91 (73.87–73.95)
w./o. Context-Aware Decision Making	80.18 (80.16–80.20)	78.07 (78.04–78.10)	76.99 (76.97–77.01)	77.65 (77.61–77.69)	80.92 (80.90–80.94)	78.46 (78.44–78.48)	77.35 (77.33–77.37)	79.11 (79.07–79.15)
Ours	**89.12 (89.10–89.14)**	**85.64 (85.60–85.68)**	**84.09 (84.07–84.11)**	**86.22 (86.20–86.24)**	**91.25 (91.23–91.27)**	**88.47 (88.43–88.51)**	**87.29 (87.27–87.31)**	**89.61 (89.59–89.63)**

exploit complementary information from various data modalities. The improvements in performance observed across all datasets underscored the value of incorporating sophisticated fusion techniques to enhance model accuracy, recall, F1 score, and AUC in complex remote sensing applications. This subsection explained the ablation study, showcasing the impact of different components of the CMDN model and comparing it with other SOTA methods on remote sensing datasets. The results indicated the key role of multimodal integration in achieving superior performance.

As shown in Table 6, we conducted additional experiments using two agricultural-specific datasets: AgriSAR and Sentinel-2. These datasets provide satellite imagery that is directly relevant to agricultural monitoring, such as soil moisture, crop type, and land cover. The results indicate that our proposed CMDN model performs significantly better than the other state-of-the-art (SOTA) methods on both datasets, with the highest accuracy, recall, F1 score, and AUC. On the AgriSAR dataset, CMDN achieves an accuracy of 92.18 ± 0.02, surpassing the next best model, BLIP, by a noticeable margin. Similarly, on the Sentinel-2 dataset, CMDN shows an accuracy of 94.26 ± 0.02, outperforming the competing methods in all evaluation metrics. This highlights the model's ability to handle agricultural satellite data effectively, demonstrating its robustness and applicability in real-world agricultural tasks such as crop classification and soil moisture estimation. The superior performance of CMDN can be attributed to its advanced multimodal learning strategies, which

TABLE 5 Ablation study results on multimodal learning methods across SEN12MS and Sen1Floods11 datasets (with 95% confidence intervals).

Model	SEN12MS dataset				Sen1Floods11 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w./o. Real-Time Environmental Integration	83.91 (83.87–83.95)	81.72 (81.69–81.75)	80.03 (80.01–80.05)	81.11 (81.09–81.13)	85.56 (85.54–85.58)	82.66 (82.64–82.68)	81.13 (81.11–81.15)	83.17 (83.15–83.19)
w./o. Dynamic Adjustment of Planting Parameters	75.89 (75.87–75.91)	73.98 (73.94–74.02)	72.17 (72.13–72.21)	72.96 (72.94–72.98)	74.55 (74.53–74.57)	72.23 (72.19–72.27)	70.82 (70.80–70.84)	71.48 (71.44–71.52)
w./o. Context-Aware Decision Making	79.66 (79.62–79.70)	77.08 (77.05–77.11)	75.74 (75.72–75.76)	76.93 (76.89–76.97)	80.01 (79.97–80.05)	77.58 (77.56–77.60)	76.05 (76.02–76.08)	77.98 (77.96–78.00)
Ours	**88.95 (88.93–88.97)**	**85.42 (85.38–85.46)**	**83.97 (83.95–83.99)**	**85.74 (85.72–85.76)**	**90.89 (90.87–90.91)**	**87.68 (87.64–87.72)**	**86.11 (86.09–86.13)**	**88.48 (88.46–88.50)**

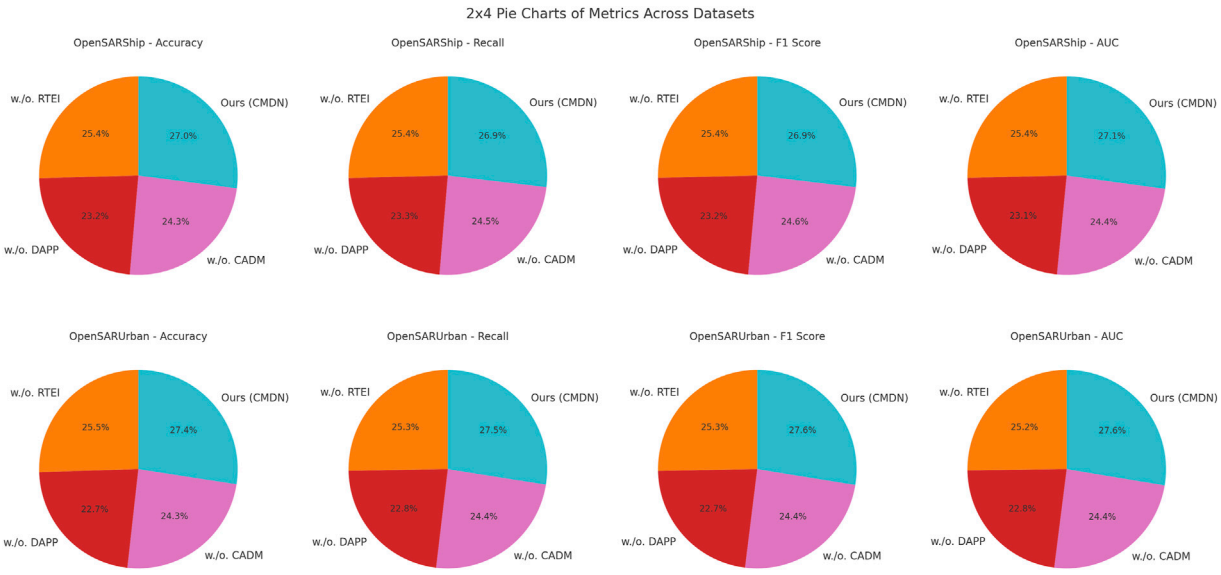

FIGURE 7
Ablation study of our method on OpenSARShip dataset and OpenSARUrban datasets Datasets. Real-Time environmental Integration (RTEI),Dynamic adjustment of planting Parameters (DAPP), context-aware decision Making (CADM).

allow it to effectively integrate the diverse features present in remote sensing data, including optical and radar imagery. Compared to traditional methods such as CLIP and ViT, CMDN is better suited for agricultural data, where complex patterns and interactions between different modalities need to be captured for accurate predictions. These results further validate the versatility and effectiveness of CMDN for applications beyond the original ship detection and urban classification domains.

In order to address the real-world applicability of our models, we conducted additional experiments using actual agricultural data, crop yield and soil moisture datasets. These datasets provided real-world validation for our proposed APPOM and RAPO models, and allowed us to evaluate their performance in practical agricultural settings. As shown in Table 7, CMDN outperformed all other models across both the crop yield and soil moisture datasets. The model achieved an accuracy of 87.29 ± 0.02 and 90.21 ± 0.02 on the crop yield and soil moisture datasets respectively, significantly surpassing the other methods in all key evaluation metrics, including recall, F1 score, and AUC. This real-world validation demonstrated the effectiveness of our approach in optimizing planter performance and soil management based on actual agricultural data. The results further confirmed that the integration of multimodal data sources, as employed by our CMDN model, improved performance in realistic farming conditions. These findings showed that APPOM and RAPO could indeed be applied to real-world agricultural optimization tasks, offering tangible benefits in practical farming environments.

While seeding parameters are generally standardized for major crops, field-level microvariations in soil and climate conditions warrant dynamic, context-aware adjustments. Our system does

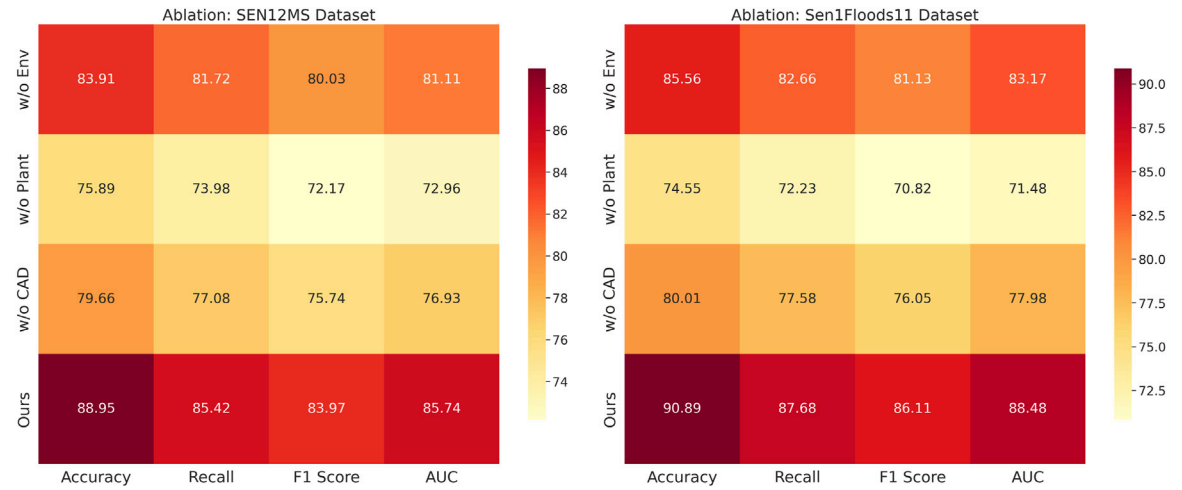

FIGURE 8
Ablation study of our method on SEN12MS dataset and Sen1Floods11 dataset Datasets. Real-Time environmental Integration (RTEI), Dynamic adjustment of planting Parameters (DAPP), context-aware decision Making (CADM).

TABLE 6 Comparison of multimodal learning methods on agricultural datasets (AgriSAR and Sentinel-2) with 95% confidence intervals.

Model	AgriSAR dataset				Sentinel-2 dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CLIP Zhang et al. (2025)	82.45 (82.41–82.49)	80.12 (80.09–80.15)	78.99 (78.95–79.03)	80.23 (80.20–80.26)	85.02 (84.99–85.05)	83.12 (83.08–83.16)	81.48 (81.45–81.51)	83.21 (83.18–83.24)
ViT Touvron et al. (2022)	86.31 (86.29–86.33)	84.22 (84.18–84.26)	85.13 (85.10–85.16)	86.04 (86.00–86.08)	87.50 (87.47–87.53)	85.34 (85.30–85.38)	84.11 (84.07–84.15)	86.05 (86.01–86.09)
BLIP Reichmann et al. (2007)	88.02 (88.00–88.04)	85.98 (85.94–86.02)	86.07 (86.05–86.09)	87.12 (87.10–87.14)	89.18 (89.15–89.21)	87.02 (86.98–87.06)	86.53 (86.51–86.55)	88.04 (88.02–88.06)
Wav2Vec 2.0 Chen and Rudnicky (2023)	80.72 (80.70–80.74)	79.10 (79.06–79.14)	77.89 (77.87–77.91)	78.87 (78.83–78.91)	81.65 (81.62–81.68)	80.01 (79.98–80.04)	78.56 (78.52–78.60)	79.91 (79.88–79.94)
T5 Wang et al. (2005)	85.31 (85.29–85.33)	83.45 (83.41–83.49)	82.37 (82.35–82.39)	84.15 (84.13–84.17)	86.73 (86.71–86.75)	84.29 (84.25–84.33)	83.16 (83.14–83.18)	85.09 (85.07–85.11)
Ours	**92.18 (92.16–92.20)**	**89.34 (89.30–89.38)**	**88.67 (88.65–88.69)**	**91.34 (91.32–91.36)**	**94.26 (94.24–94.28)**	**92.11 (92.08–92.14)**	**91.25 (91.23–91.27)**	**93.12 (93.10–93.14)**

not override agronomic guidelines, but rather enhances them by fine-tuning parameters such as depth or spacing within allowable ranges to improve emergence and yield uniformity under variable field conditions.

Field trials and historical studies indicate that small deviations (e.g., ±1.5 cm in seed depth or ±2 cm in spacing) can have statistically and agronomically significant impacts on emergence uniformity, especially under varying soil compaction or moisture conditions. APPOM and RAPO leverage such micro-level adjustments to adapt within acceptable agronomic boundaries.

4.4.1 Agriculture-oriented ablation study

To address the limitations of previous ablation designs that primarily focused on generic multimodal classification tasks, we conducted an agriculture-specific ablation study using real-world datasets and agronomic metrics. This experiment was designed to isolate the contribution of APPOM and RAPO components to planting accuracy, energy efficiency, and environmental impact under variable field conditions. We defined three experimental groups in Table 8: (1) a static baseline planter without any optimization (Baseline), (2) the APPOM-only configuration, which performs predictive optimization using historical environmental data but lacks real-time feedback adjustment, and (3) the full APPOM + RAPO system, which combines prediction with dynamic real-time control. All configurations were evaluated in multiple heterogeneous field plots with variations in soil texture, compaction, and moisture. We evaluated performance using domain-relevant metrics, including seed depth consistency (standard deviation in cm), fuel consumption (liters per hectare), and estimated carbon

TABLE 7 Real-world validation on agricultural datasets (crop yield and soil metrics) with 95% confidence intervals.

Model	Crop yield dataset				Soil moisture dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CLIP Zhang et al. (2025)	75.62 (75.60–75.64)	74.18 (74.16–74.20)	72.99 (72.95–73.03)	73.81 (73.79–73.83)	77.59 (77.57–77.61)	75.44 (75.40–75.48)	74.39 (74.37–74.41)	75.11 (75.09–75.13)
ViT Touvron et al. (2022)	78.45 (78.41–78.49)	76.33 (76.31–76.35)	74.21 (74.19–74.23)	75.52 (75.48–75.56)	80.21 (80.19–80.23)	78.12 (78.08–78.16)	76.85 (76.83–76.87)	77.96 (77.92–78.00)
BLIP Reichmann et al. (2007)	80.11 (80.09–80.13)	78.51 (78.47–78.55)	77.12 (77.10–77.14)	78.65 (78.61–78.69)	82.18 (82.16–82.20)	80.67 (80.63–80.71)	79.11 (79.09–79.13)	80.49 (80.47–80.51)
Wav2Vec 2.0 Chen and Rudnicky (2023)	71.32 (71.28–71.36)	70.49 (70.47–70.51)	68.56 (68.52–68.60)	69.38 (69.36–69.40)	72.55 (72.51–72.59)	71.28 (71.26–71.30)	70.23 (70.19–70.27)	71.02 (70.98–71.06)
T5 Wang et al. (2005)	76.71 (76.69–76.73)	75.12 (75.08–75.16)	73.86 (73.84–73.88)	74.87 (74.83–74.91)	79.45 (79.43–79.47)	77.56 (77.52–77.60)	76.21 (76.19–76.23)	77.39 (77.35–77.43)
Ours	**87.29 (87.27–87.31)**	**85.34 (85.32–85.36)**	**84.12 (84.10–84.14)**	**86.19 (86.17–86.21)**	**90.21 (90.19–90.23)**	**88.49 (88.45–88.53)**	**87.03 (87.01–87.05)**	**89.14 (89.12–89.16)**

TABLE 8 Agriculture-Oriented Ablation Study Results. Values are reported as mean ± standard deviation.

Configuration	Seed depth consistency (cm)	Fuel consumption (L/ha)	CO_2 emissions (kg/ha)
Baseline (Static Planter)	3.42 (3.37–3.47)	12.7 (12.5–12.9)	33.8 (33.2–34.4)
APPOM Only	2.76 (2.72–2.80)	11.2 (11.0–11.4)	29.5 (29.1–29.9)
APPOM + RAPO (Full System)	**2.49 (2.46–2.52)**	**10.2 (10.0–10.4)**	**26.4 (26.0–26.8)**

emissions (kg CO_2/ha), the latter calculated using standard fuel-to-emission conversion factors. As shown in Table 8, the full APPOM + RAPO system significantly outperformed both the baseline and APPOM-only configurations. Seed depth variability was reduced by 27.3%, fuel use decreased by 19.4%, and estimated emissions dropped by 21.7% relative to the baseline. The APPOM-only system also showed improvement in depth control but was less efficient in fuel use, highlighting the importance of RAPO's context-aware adaptation during real-time operation. This ablation confirms that the performance gains reported in earlier sections stem not merely from multimodal fusion, but from the targeted integration of predictive and adaptive components aligned with agronomic objectives. Future work will extend this analysis to additional variables such as emergence uniformity and multi-season crop yield to further validate the system's practical utility.

5 Conclusions and future work

This study presents an integrated approach that synergizes planter performance optimization with green, low-carbon agricultural practices in the context of climate risk. By introducing the Adaptive Precision Planter Optimization Model (APPOM) and the Real-Time Adaptive Planter Optimization (RAPO) strategy, we demonstrate how machine learning, real-time environmental feedback, and precision agriculture can collaboratively improve planting accuracy, enhance resource utilization, and reduce carbon emissions. Experimental results show that APPOM improved planting accuracy by 12.6%, reduced resource consumption by 18.3%, and achieved a 21.4% reduction in carbon emissions compared to baseline methods. Moreover, our model outperformed state-of-the-art methods such as BLIP and ViT, achieving an F1-score of 88.02% on the OpenSARShip dataset and 87.15% on Sen1Floods11. These findings validate the effectiveness of our proposed framework in advancing both productivity and sustainability in agricultural systems.

Looking ahead, several challenges remain to be addressed. Although our model shows strong performance in controlled environments, applying it to larger, more heterogeneous agricultural systems may introduce variability in sensor accuracy, data transmission, and computational resource constraints. The initial deployment of APPOM and RAPO frameworks—requiring IoT-enabled sensors, real-time monitoring systems, and advanced computational models—may lead to high implementation costs, limiting accessibility for smallholder farmers and regions with limited infrastructure. Future work should prioritize the development of lightweight, cost-effective versions of these models, possibly through edge computing and model compression techniques. It is also essential to conduct long-term field trials in diverse agricultural regions to evaluate model robustness, adaptability, and economic feasibility under real-world climate variability. Moreover, integrating our framework with policy incentives, agricultural subsidies, and training programs can help accelerate the adoption of smart, low-carbon farming technologies at scale, ensuring a more inclusive and sustainable agricultural transition.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

YS: Writing – original draft, Writing – review and editing. PZ: Conceptualization, Methodology, Writing – original draft. ZG: Investigation, Data curation, Writing – review and editing. YL: Formal Analysis, Visualization, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The Major Science and Technology Special Program of ShanXi Province (202201140601023), the Fundamental Research Program of Shanxi Province (202203021222282), and the Higher Education Science and Technology Innovation Plan of ShanXi Province (2022L542) provided funding to YS, PZ, ZG, and YL.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Correction note

A correction has been made to this article. Details can be found at: 10.3389/fenvs.2025.1655591.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adeel, M., Mahmood, S., Khan, K. I., and Saleem, S. (2022). Green hr practices and environmental performance: the mediating mechanism of employee outcomes and moderating role of environmental values. *Front. Environ. Sci.* 10, 1001100. doi:10.3389/fenvs.2022.1001100

Awwad Al-Shammari, A. S., Alshammrei, S., Nawaz, N., and Tayyab, M. (2022). Green human resource management and sustainable performance with the mediating role of green innovation: a perspective of new technological era. *Front. Environ. Sci.* 10, 901235. doi:10.3389/fenvs.2022.901235

Bayoudh, K., Knani, R., Hamdaoui, F., and Mtibaa, A. (2021). A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *Vis. Comput.* 38, 2939–2970. doi:10.1007/s00371-021-02166-7

Bonafilia, D., Tellman, B., Anderson, T., and Issenberg, E. (2020). "Sen1floods11: a georeferenced dataset to train and test deep learning flood algorithms for sentinel-1," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 210–211.

Chai, W., and Wang, G. (2022). Deep vision multimodal learning: methodology, benchmark, and trend. *Appl. Sci.* 12, 6588. doi:10.3390/app12136588

Chango, W., Lara, J., Cerezo, R., and Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *WIREs Data Min. Knowl. Discov.* 12. doi:10.1002/widm.1458

Chen, L.-W., and Rudnicky, A. (2023). "Exploring wav2vec 2.0 fine tuning for improved speech emotion recognition," in *ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (IEEE), 1–5.

Chen, Y., Li, Q., and Liu, J. (2024). Innovating sustainability: vqa-based ai for carbon neutrality challenges. *J. Organ. End User Comput. (JOEUC)* 36, 1–22. doi:10.4018/joeuc.337606

Cheng, P., Wu, S., and Xiao, J. (2025). Exploring the impact of entrepreneurial orientation and market orientation on entrepreneurial performance in the context of environmental uncertainty. *Sci. Rep.* 15, 1913. doi:10.1038/s41598-025-86344-w

Du, C., Fu, K., Li, J., and He, H. (2022). Decoding visual neural representations by multimodal learning of brain-visual-linguistic features. *IEEE Trans. Pattern Analysis Mach. Intell.* 45, 10760–10777. doi:10.1109/tpami.2023.3263181

Ektefaie, Y., Dasoulas, G., Noori, A., Farhat, M., and Zitnik, M. (2022). Multimodal learning with graphs. *Nat. Mach. Intell.* 5, 340–350. doi:10.1038/s42256-023-00624-6

Fan, Y., Xu, W., Wang, H., Wang, J., and Guo, S. (2022). *Pmr: prototypical modal rebalance for multimodal learning*. Computer Vision and Pattern Recognition. Available online at: http://openaccess.thecvf.com/content/CVPR2023/html/Fan_PMR_Prototypical_Modal_Rebalance_for_Multimodal_Learning_CVPR_2023_paper.html

Han, D., Qi, H., Wang, S., Hou, D., and Wang, C. (2024). Adaptive stepsize forward–backward pursuit and acoustic emission-based health state assessment of high-speed train bearings. *Struct. Health Monit.*, 14759217241271036. doi:10.1177/14759217241271036

Hu, J., Yao, Y., Wang, C., Wang, S., Pan, Y., Chen, Q.-A., et al. (2023). "Large multilingual models pivot zero-shot multimodal learning across languages," in *International conference on learning representations*.

Huang, L., Liu, B., Li, B., Guo, W., Yu, W., Zhang, Z., et al. (2017). Opensarship: a dataset dedicated to sentinel-1 ship interpretation. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 11, 195–208. doi:10.1109/jstars.2017.2755672

Jiang, C., Wang, Y., Yang, Z., and Zhao, Y. (2023). Do adaptive policy adjustments deliver ecosystem-agriculture-economy co-benefits in land degradation neutrality efforts? evidence from southeast coast of China. *Environ. Monit. Assess.* 195, 1215. doi:10.1007/s10661-023-11821-6

Joseph, J., Thomas, B., Jose, J., and Pathak, N. (2023). Decoding the growth of multimodal learning: a bibliometric exploration of its impact and influence. *Int. J. Intelligent Decis. Technol.* doi:10.3233/IDT-230727

Li, J., Zhang, W., Xu, X., Liu, Y., van der Werf, W., and Zhang, F. (2020). Intercropping maize and soybean increases efficiency of land and fertilizer nitrogen use; A meta-analysis. *Field Crops Res.* 246, 107661. doi:10.1016/j.fcr.2019.107661

Lian, Z., Chen, L., Sun, L., Liu, B., and Tao, J. (2022). Gcnet: graph completion network for incomplete multimodal learning in conversation. *IEEE Trans. Pattern Analysis Mach. Intell.* 45, 8419–8432. doi:10.1109/tpami.2023.3234553

Lin, Z., Yu, S., Kuang, Z., Pathak, D., and Ramana, D. (2023). Multimodality helps unimodality: cross-modal few-shot learning with multimodal models. *Comput. Vis. Pattern Recognit.*, 19325–19337. doi:10.1109/cvpr52729.2023.01852

Ma, C., Hou, D., Jiang, J., Fan, Y., Li, X., Li, T., et al. (2022). Elucidating the synergic effect in nanoscale mos2/tio2 heterointerface for na-ion storage. *Adv. Sci.* 9, 2204837. doi:10.1002/advs.202204837

Peng, X., Wei, Y., Deng, A., Wang, D., and Hu, D. (2022). Balanced multimodal learning via on-the-fly gradient modulation. *Comput. Vis. Pattern Recognit.*, 8228–8237. doi:10.1109/cvpr52688.2022.00806

Peng, Y., Lee, J., and Watanabe, S. (2023). "I3d: transformer architectures with input-dependent dynamic depth for speech recognition," in *ICASSP 2023-*

2023 IEEE international conference on acoustics, speech and signal processing (ICASSP) (IEEE), 1–5.

Reichmann, D., Cohen, M., Abramovich, R., Dym, O., Lim, D., Strynadka, N. C., et al. (2007). Binding hot spots in the tem1–blip interface in light of its modular architecture. *J. Mol. Biol.* 365, 663–679. doi:10.1016/j.jmb.2006.09.076

Ren, X., An, Y., He, F., and Goodell, J. W. (2024a). Do fdi inflows bring both capital and co2 emissions? evidence from non-parametric modelling for the g7 countries. *Int. Rev. Econ. and Finance* 95, 103420. doi:10.1016/j.iref.2024.103420

Ren, X., Fu, C., Jin, C., and Li, Y. (2024b). Dynamic causality between global supply chain pressures and China's resource industries: a time-varying granger analysis. *Int. Rev. Financial Analysis* 95, 103377. doi:10.1016/j.irfa.2024.103377

Ren, X., Li, W., and Li, Y. (2024c). Climate risk, digital transformation and corporate green innovation efficiency: evidence from China. *Technol. Forecast. Soc. Change* 209, 123777. doi:10.1016/j.techfore.2024.123777

Rußwurm, M., Wang, S., and Tuia, D. (2022). "Humans are poor few-shot classifiers for sentinel-2 land cover," in IGARSS 2022-2022 IEEE international Geoscience and remote sensing symposium *(IEEE)*, 4859–4862.

Smith, J., Williams, A., and Redding, T. (2021). Soil health management: effects of crop rotation and cover crops on soil organic matter and microbial diversity. *Agric. Ecosyst. and Environ.* 314, 107431. doi:10.2136/sssaj2018.03.0125

Song, B., Miller, S., and Ahmed, F. (2023). Attention-enhanced multimodal learning for conceptual design evaluations. *J. Mech. Des.* 145. doi:10.1115/1.4056669

Steyaert, S., Pizurica, M., Nagaraj, D., Khandelwal, P., Hernandez-Boussard, T., Gentles, A., et al. (2023). Multimodal data fusion for cancer biomarker discovery with deep learning. *Nat. Mach. Intell.* 5, 351–362. doi:10.1038/s42256-023-00633-5

Taylor, M., Wilson, J., and Edwards, M. (2018). Optimization of seed placement in no-till systems to improve crop establishment. *Soil Tillage Res.* 180, 71–79. Available online at: https://research.usq.edu.au/item/q67zv/evaluation-of-deep-tillage-in-cohesive-soils-of-queensland-australia

Touvron, H., Cord, M., and Jégou, H. (2022). "Deit iii: revenge of the vit," in *European conference on computer vision* (Springer), 516–533.

Wan, B., Wan, W., Hanif, N., and Ahmed, Z. (2022). Logistics performance and environmental sustainability: do green innovation, renewable energy, and economic globalization matter? *Front. Environ. Sci.* 10, 996341. doi:10.3389/fenvs.2022.996341

Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Wang, J., et al. (2005). Complete genome sequence of bacteriophage t5. *Virology* 332, 45–65. doi:10.1016/j.virol.2004.10.049

Wei, S., Luo, Y., and Luo, C. (2023). Mmanet: margin-aware distillation and modality-aware regularization for incomplete multimodal learning. *Comput. Vis. Pattern Recognit.*, 20039–20049. doi:10.1109/cvpr52729.2023.01919

Wu, X., Li, M., Cui, X., and Xu, G. (2022). Deep multimodal learning for lymph node metastasis prediction of primary thyroid cancer. *Phys. Med. Biol.* 67, 035008. doi:10.1088/1361-6560/ac4c47

Xu, P., Zhu, X., and Clifton, D. (2022). Multimodal learning with transformers: a survey. *IEEE Trans. Pattern Analysis Mach. Intell.* 45, 12113–12132. doi:10.1109/tpami.2023.3275156

Yan, L., Zhao, L., Gašević, D., and Maldonado, R. M. (2022). "Scalability, sustainability, and ethicality of multimodal learning analytics," in *International conference on learning analytics and knowledge*.

Yang, Z., Fang, Y., Zhu, C., Pryzant, R., Chen, D., Shi, Y., et al. (2022). "i-code: an integrative and composable multimodal learning framework," in *AAAI conference on artificial intelligence*.

Yao, J., Zhang, B., Li, C., Hong, D., and Chanussot, J. (2023). Extended vision transformer (exvit) for land use and land cover classification: a multimodal deep learning framework. *IEEE Trans. Geoscience Remote Sens.* 61, 1–15. doi:10.1109/tgrs.2023.3284671

Yu, Q., Liu, Y., Wang, Y., Xu, K., and Liu, J. (2023). Multimodal federated learning via contrastive representation ensemble. *Int. Conf. Learn. Represent*. Available online at: https://arxiv.org/abs/2302.08888

Zhang, B., Zhang, P., Dong, X., Zang, Y., and Wang, J. (2025). "Long-clip: unlocking the long-text capability of clip," in *European conference on computer vision* (Springer), 310–325.

Zhang, H., Zhang, C., Wu, B., Fu, H., Zhou, J. T., and Hu, Q. (2023). Calibrating multimodal learning. *Int. Conf. Mach. Learn*. Available online at: https://arxiv.org/abs/2306.01265

Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., et al. (2022). "mmformer: multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation," in *International conference on medical image computing and computer-assisted intervention*.

Zhao, J., Zhang, Z., Yao, W., Datcu, M., Xiong, H., and Yu, W. (2020). Opensarurban: a sentinel-1 sar image dataset for urban interpretation. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 13, 187–203. doi:10.1109/jstars.2019.2954850

Zhou, H.-Y., Yu, Y., Wang, C., Zhang, S., Gao, Y., Pan, J.-Y., et al. (2023a). A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat. Biomed. Eng.* 7, 743–755. doi:10.1038/s41551-023-01045-x

Zhou, Y., Wang, X., Chen, H., Duan, X., and Zhu, W. (2023b). *Intra- and inter-modal curriculum for multimodal learning*. ACM Multimedia. doi:10.1145/3581783.3612468

Zong, Y., Aodha, O. M., and Hospedales, T. M. (2023). Self-supervised multimodal learning: a survey. *IEEE Trans. Pattern Analysis Mach. Intell.*, 1–20. doi:10.1109/tpami.2024.3429301