USOD PUBLISHERS
United Scientific Open Discovery

Journal of Advanced Artificial Intelligence Systems and Frontiers
Volume 1 | Issue 2

Research Article

Beyond Task-Specific Models: Toward PALM-FM, a Multimodal Spatiotemporal Foundation Model for Generalizable Oil-Palm Plantation Intelligence under Tropical Domain Shift

Loso Judijanto

IPOSS Jakarta, Indonesia

***Corresponding author:** Loso Judijanto IPOSS Jakarta, Indonesia.

Citation: Loso Judijanto (2026) Beyond Task-Specific Models: Toward PALM-FM, a Multimodal Spatiotemporal Foundation Model for Generalizable Oil-Palm Plantation Intelligence under Tropical Domain Shift. J of Adv Art Int Sys & Frontiers 1(2):1-14.

Received Date: October 07, 2026
Accepted Date: October 12, 2026
Published Date: November 11, 2026

Abstract

Artificial intelligence has become increasingly embedded in oil-palm plantation management, yet most applications remain task-specific, geographically constrained, and dependent on data collected from particular estates, seasons, sensors, or management systems. This qualitative literature review examines whether the next phase of oil-palm artificial intelligence should shift from isolated predictive models toward an oil-palm-specific multimodal spatiotemporal foundation model. Recent literature on oil-palm remote sensing, agricultural artificial intelligence, self-supervised representation learning, multimodal fusion, domain adaptation, geospatial foundation models, and responsible AI is synthesized thematically. The review identifies tropical domain shift as a central limitation to operational generalization. Geographic heterogeneity, plantation age, climate variability, cloud conditions, sensor differences, soil characteristics, management practices, and contrasts between industrial and smallholder systems can substantially alter data distributions. To address these limitations, the article proposes PALM-FM, a conceptual architecture integrating multimodal plantation data, spatiotemporal self-supervised pretraining, cross-modal alignment, parameter-efficient downstream adaptation, uncertainty estimation, and cross-domain benchmarking. PALM-FM could support plantation mapping, tree inventory, age and height estimation, disease surveillance, nutrient assessment, yield forecasting, fruit maturity evaluation, replanting monitoring, and sustainability verification through a reusable representation. A phased research roadmap is proposed to transition oil-palm AI from fragmented models toward generalizable, responsible, and continually improving plantation intelligence.

Keywords: oil palm; foundation model; multimodal artificial intelligence; spatiotemporal learning; remote sensing; domain shift; self-supervised learning; precision plantation management; transfer learning; tropical agriculture

Introduction

Digital Transformation and the Expansion of Oil-Palm Artificial Intelligence

Digital agriculture has progressed from the use of isolated sensors and decision-support tools toward interconnected systems combining remote sensing, artificial intelligence, cloud computing, geospatial analytics, mobile platforms, robotics, and farm-management information. Agriculture 4.0 reflects this transition by enabling observations to be collected continuously, integrated across spatial scales, and converted into operational recommendations through increasingly sophisticated computational models. Nevertheless, the literature also shows that interoperability, data heterogeneity, deployment costs, and technological fragmentation continue to constrain the practical integration of these technologies (da Silveira et al., 2021; Li et al., 2024).

Oil-palm production provides an especially compelling setting for such digital transformation. Plantations occupy large and geographically dispersed areas, the crop remains productive for decades, individual palms can be geolocated, management activities generate large amounts of structured operational information, and yield is influenced by interactions among weather, soil, age, management, biological stress, and historical interventions. Consequently, conventional field scouting alone is increasingly insufficient for plantation systems that require timely information at palm, block, estate, and regional scales.

Remote sensing and artificial intelligence have already generated substantial advances in oil-palm observation. Machine-learning models have been used to map plantations across large territories, identify individual palms, estimate density, reconstruct plantation age, predict yield, classify fruit ripeness, assess nutrients, and detect disease symptoms (Shaharum et al., 2020; Rodríguez et al., 2021; Wibowo et al., 2022; Zang et al., 2023). At regional scale, satellite imagery enables plantation monitoring over areas that would be prohibitively expensive to survey continuously in the field. At estate and palm scales, unmanned aerial vehicles and computer vision enable increasingly detailed inventory and canopy observations (Lee et al., 2024).

Artificial intelligence has also expanded beyond geometric recognition toward biological and operational interpretation. Hyperspectral sensing has been applied to macronutrient assessment and disease detection, while machine-learning and deep-learning approaches have increasingly incorporated weather, soil moisture, agronomic variables, and production histories into yield forecasting (Amirruddin et al., 2020; Lee et al., 2022; Khan et al., 2022; Ang et al., 2022). Fruit-level systems have moved toward mobile and video-based applications capable of supporting harvesting and quality decisions close to the point of operation (Suharjito et al., 2021; Junior and Suharjito, 2023).

These developments demonstrate increasing technical maturity. However, they also reveal a structural limitation: oil-palm AI is primarily organized as a collection of task-specific systems rather than as a unified representation of plantation intelligence. A model developed for palm counting usually learns independently from one developed for Ganoderma detection. Yield models normally do not inherit representations from plantation mapping models, and fruit-maturity systems are commonly separated from broader plantation-state estimation. The field therefore accumulates algorithms faster than it accumulates reusable representations of the underlying biological production system.

From Task-Specific Models to Foundation Models

Foundation models provide a fundamentally different development paradigm. Instead of training a complete model independently for every downstream objective, large and diverse datasets are used for pretraining so that reusable representations can later be adapted to multiple tasks. In agriculture, this approach is particularly attractive because expert annotation remains expensive and unevenly distributed, while enormous volumes of unlabelled satellite images, UAV observations, field photographs, sensor streams, and farm records are becoming available (Li et al., 2024; Yin et al., 2025).

Recent agricultural research indicates that foundation-model principles are becoming increasingly practical. Foundation vision models can be adapted for disease, weed, and growth-stage recognition through linear probing, parameter-efficient methods such as low-rank adaptation, and knowledge distillation rather than complete retraining (Espejo-Garcia et al., 2025). MetaFruit demonstrates how heterogeneous fruit datasets can be organized to support broader agricultural foundation-model development rather than narrowly bounded classifiers (Li et al., 2025). These developments indicate that agricultural artificial intelligence may be entering a stage in which representation reuse becomes as important as architecture-specific accuracy.

The same transition is occurring in Earth observation. Remote-sensing foundation models increasingly use massive satellite archives, masked modelling, self-supervised learning, temporal representation, and multimodal alignment to create general-purpose geospatial features (Xiao et al., 2025). Recent perspectives and surveys frame these models as a route toward generalist Earth-observation intelligence while emphasizing unresolved challenges in transferability, modality alignment, and benchmark design (Zhang et al., 2024; Huo et al., 2025; Zhou et al., 2025).

Agricultural examples published during 2025–2026 make this direction increasingly concrete. CropSTS applies decoupled spatiotemporal attention and self-supervised strategies to fine-grained agricultural classification, recognizing that crop intelligence requires stronger temporal encoding than many generic geospatial foundation models provide (Yan et al., 2025). AgriFM uses multi-source temporal observations and a globally representative pretraining corpus exceeding 25 million samples to support several agricultural mapping tasks through a common framework (Li et al., 2026a). SeaMo similarly incorporates multimodality and seasonal information into remote-sensing foundation-model pretraining, demonstrating the importance of learning temporal and cross-modal relationships jointly rather than treating seasonality as an auxiliary input (Li et al., 2026b).

These developments create a strong conceptual opportunity for oil palm. Instead of asking how to build increasingly specialized models for each plantation task, a more fundamental question is whether the crop can be represented through a shared multimodal and spatiotemporal foundation model that is subsequently adapted to multiple plantation applications.

Tropical Domain Shift as the Core Generalization Challenge

A foundation model becomes useful only when its learned representation remains informative beyond its immediate training environment. This consideration is especially important for oil palm because plantations occur across heterogeneous tropical landscapes characterized by varying rainfall, cloud cover, topography, soils, planting materials, palm ages, disease environments, management systems, imaging conditions, and sensor infrastructures.

This heterogeneity creates domain shift. A statistical relationship learned from one estate does not automatically remain stable in another estate. A model trained using one satellite or UAV camera may encounter different resolution, spectral response, noise, or viewing geometry after deployment with another sensor. Likewise, imagery collected from mature plantations may not adequately represent recently planted or aging palms.

Remote-sensing research demonstrates that such distribution changes can materially reduce machine-learning performance. Domain-adaptation reviews show that geographic, temporal, sensor, and semantic shifts are persistent obstacles to transferring classifiers between environments (Peng et al., 2022). Crop-mapping studies demonstrate that models trained in a particular place and year can deteriorate when applied across regions or seasons because spectral and phenological distributions change (Wang et al., 2023). Cross-sensor research further shows that differences among acquisition systems may require explicit adaptation rather than assuming sensor interchangeability (Wang et al., 2022).

Temporal displacement is particularly important in agriculture. TimeMatch demonstrates that estimating temporal shifts can improve cross-region crop classification where phenological stages are not synchronized (Nyborg et al., 2022). Unsupervised adversarial adaptation and source-free adaptation provide alternative strategies when labelled data from the target environment are scarce or when source datasets cannot be redistributed (Kwak and Park, 2022; Mohammadi et al., 2024). Spatial and temporal transferability therefore needs to be treated as a principal research outcome rather than a secondary validation exercise.

Oil-palm mapping offers direct evidence of this challenge. Recent regional-scale mapping of Malaysia and Indonesia demonstrates that historical low-resolution supervision can be used to derive later high-resolution plantation maps, but temporal discrepancies between reference information and later landscape conditions remain a critical modelling problem (Kuapanich et al., 2026). Long-term stand-age mapping similarly requires decades of observations because a contemporary image alone cannot reconstruct the full plantation lifecycle (Weerakitikul et al., 2025).

This article therefore introduces **tropical domain shift** as a unifying concept describing interacting distribution changes associated with geography, climate, seasonality, plantation age, soil and terrain, sensor characteristics, management practices, biological stress, and production systems. The central proposition is that an oil-palm foundation model should be evaluated primarily through its capacity to retain useful representations under these shifts.

Research Questions and Objectives

The review is organized around seven research questions. The first

examines how the task-specific architecture of contemporary oil-palm artificial intelligence constrains representation reuse, scaling, and generalization. The second asks which forms of tropical domain shift are most consequential to transferable plantation intelligence. The third evaluates which combinations of optical, radar, hyperspectral, LiDAR, UAV, weather, soil, agronomic, operational, and textual data should constitute the multimodal information architecture of an oil-palm foundation model.

The fourth research question considers which self-supervised, masked-modelling, contrastive, temporal, and multimodal learning mechanisms are most appropriate for oil-palm pretraining. The fifth asks how a common pretrained representation could be adapted to multiple downstream tasks, including plantation mapping, inventory, disease surveillance, nutrient assessment, age and height estimation, yield forecasting, and fruit maturity analysis. The sixth asks how benchmarking should be redesigned to distinguish genuine cross-domain generalization from apparently high accuracy obtained from within-domain random test sets. Finally, the seventh evaluates the governance, privacy, explainability, uncertainty, and stakeholder arrangements needed for responsible deployment.

The overarching objective is to synthesize recent evidence and develop a conceptual architecture for an oil-palm-specific foundation model. The review does not attempt to identify a universally superior algorithm. Instead, it seeks to define the technical and institutional components required for reusable and transferable plantation intelligence.

Literature Review

Conceptual Foundations of Agricultural Foundation Models

Conventional supervised artificial intelligence generally optimizes model parameters for a predefined labelled task. This architecture has produced important agricultural applications but requires substantial annotation effort and commonly embeds assumptions specific to the original dataset. Support vector machines remain effective for many precision-agriculture applications, particularly when sample sizes are moderate and feature representations are carefully constructed, yet the broader trajectory of agricultural analytics increasingly favours models capable of learning representations directly from high-dimensional inputs (Kok et al., 2021).

Plant-disease recognition illustrates both the power and limitations of conventional deep learning. Convolutional neural networks can achieve high classification performance, but deployment outside controlled datasets is complicated by environmental variation, class imbalance, imaging differences, and limited training diversity (Abade et al., 2021). Similar difficulties occur in hyperspectral crop mapping, where architecture performance depends strongly on sensor conditions, label availability, and spatial–spectral generalization (Meng et al., 2021).

Foundation models seek to reduce this dependence on narrowly labelled tasks. They are pretrained on large datasets using objectives that encourage the emergence of generalizable representations. Adaptation can then be performed through full fine-tuning, linear probing, lightweight adapters, prompt-like mechanisms, or parameter-efficient methods.

Agricultural foundation-model reviews identify data heterogeneity, limited labels, distribution shift, training cost, deployment efficiency, and multimodal integration as major challenges (Li et al., 2024; Yin et al., 2025). Importantly, agricultural foundation models cannot simply replicate internet-scale vision models because agricultural data are strongly temporal, environmentally conditioned, spatially structured, and frequently collected through specialized sensors. This makes domain-specific pretraining and agronomic inductive biases particularly important.

Self-Supervised Representation Learning

The abundance of unlabelled data makes self-supervised learning particularly relevant to plantation systems. Satellite missions continuously generate imagery, drones repeatedly survey estates, and operational databases accumulate weather, harvesting, yield, fertilization, and field observations even when expert annotations are unavailable.

Masked autoencoding is one important approach. SpectralMAE applies spectral masking to hyperspectral remote-sensing reconstruction, encouraging the model to capture relationships among spectral bands rather than memorize task labels (Zhu et al., 2023). Feature-guided masked autoencoding similarly demonstrates that remote-sensing representations can be learned from spatial information before downstream supervision is introduced (Wang et al., 2024). SpectralGPT further shows that large-scale spectral pretraining can accommodate varying image sizes, resolutions, time series, and geographic regions within a common representation framework (Hong et al., 2024).

Temporal reconstruction provides another opportunity. Liu et al. (2024) use self-supervised learning to reconstruct seamless harmonized Landsat–Sentinel time series, illustrating how missing or irregular Earth-observation data can be treated as a representation-learning problem. This principle is directly relevant to tropical agriculture, where persistent cloud cover creates irregular optical observations.

Contrastive learning provides a complementary strategy. Multimodal supervised contrastive learning can align information from different remote-sensing modalities while improving downstream discrimination (Berg et al., 2024). Remote-sensing foundation-model research using intra-instance similarity further demonstrates that temporal variations can expand the positive representation space and strengthen transfer across downstream tasks (An et al., 2024). For oil palm, different observations of the same block, palm, or plantation period could therefore be encouraged to occupy nearby regions in the latent representation even when one observation is optical and another is radar, UAV, or agronomic.

Multimodal Learning

Agricultural systems are intrinsically multimodal because different sensing systems observe different dimensions of the production process. Optical imagery characterizes surface reflectance and canopy appearance. Synthetic aperture radar is sensitive to structural and moisture-related properties and remains operational under clouds. Hyperspectral imagery can resolve subtle biochemical signatures, while LiDAR provides detailed vertical structure. UAVs add high spatial resolution, and field databases contribute soil, fertilizer, weather, disease, harvest, and management context.

Image-fusion research has demonstrated that agricultural monitoring can benefit from combining complementary sources rather than relying on a single sensor (Li et al., 2021). However, conventional data fusion often combines modalities only after features are independently generated. Foundation-model research instead attempts to learn correspondences among modalities during pretraining.

Cross-modal alignment is a major research challenge because modalities differ in spatial resolution, measurement physics, temporal availability, and semantic meaning. Zavras et al. (2025) describe this problem as a modality gap within remote-sensing vision–language learning. Multimodal foundation-model reviews similarly identify robust alignment as essential to transferring information across sensor types and downstream tasks (Zhou et al., 2025).

SeaMo illustrates one pathway by jointly addressing multimodality and seasonal structure (Li et al., 2026b). For oil palm, analogous learning could align optical reflectance with radar backscatter, canopy height, rainfall histories, and field-management records. Such alignment would produce a latent representation that is more biologically and operationally meaningful than any single observation stream.

Oil-Palm Mapping and Inventory

Oil-palm mapping is among the most mature AI-enabled plantation applications. Shaharum et al. (2020) demonstrate large-scale plantation mapping using Google Earth Engine and machine-learning algorithms. Mohd Najib et al. (2020) show that active and passive remote sensing can be combined effectively, underscoring the complementary value of optical and radar information in tropical environments.

Xu et al. (2021) further demonstrate machine-learning approaches for detecting oil-palm plantations from remote-sensing data. Rodríguez et al. (2021) extend the problem from plantation presence toward density estimation at country scale and show that active learning can reduce the burden of reference-data collection. At finer spatial scales, high-resolution imagery enables individual palm detection. Wibowo et al. (2022) demonstrate large-scale oil-palm tree detection using deep learning, while Lee et al. (2024) improve UAV-based detection through an enhanced RetinaNet architecture. These studies establish tree location and canopy geometry as important reusable representation elements.

Mapping in geographically distinct environments also exposes generalization issues. Abramowitz et al. (2023) examine differentiation between oil-palm plantations and natural forests in Ghana, emphasizing that landscape context differs substantially from Southeast Asian industrial plantation systems. Such variation supports the need for foundation-model pretraining that includes diverse tropical oil-palm geographies rather than treating one region as universally representative.

Plantation Age, Structure, and Temporal Dynamics

Oil palm is a perennial production system whose characteristics change over decades. Stand age influences palm height, canopy size, planting geometry, understory visibility, yield potential, labour requirements, and replanting decisions. Consequently, temporal representation is not merely beneficial but fundamental. Effendi et al. (2022) identify phenological modelling as an important frontier for oil-palm remote sensing, while Zang et al. (2023) combine stand-age information with ICESat-2 measurements to map oil-palm height. These studies show that structural characteristics are tightly linked to temporal development.

Long historical archives further enable reconstruction of plantation lifecycles. Weerakitikul et al. (2025) integrate Landsat and Sentinel time series over several decades for plantation mapping and stand-age detection. Kuapanich et al. (2026) generate 10 m time-series oil-palm mapping for Indonesia and Malaysia using Sentinel-2 and noise-robust learning from historical labels. Together, these studies demonstrate that temporal continuity is essential for distinguishing persistent plantation characteristics from transient spectral conditions.

Agricultural foundation models increasingly incorporate this principle explicitly. CropSTS introduces decoupled spatiotemporal attention tailored to cropland dynamics (Yan et al., 2025), while AgriFM jointly learns multiscale spatial and temporal agricultural representations from several satellite sources (Li et al., 2026a). PALM-FM should therefore model a plantation as a temporal trajectory rather than as a collection of independent images.

Disease, Nutrient Status, and Physiological Intelligence

Disease and nutrient applications provide a strong test of whether a foundation representation captures biological condition rather than merely geometric appearance. Amirruddin et al. (2020) use hyperspectral spectroscopy to classify oil-palm macronutrient status, illustrating the value of narrow spectral information for physiological assessment.

Ganoderma boninense and basal stem rot are particularly important disease targets. Azmi et al. (2020) show that machine learning can support early Ganoderma detection in oil-palm seedlings. Lee et al. (2022) develop hyperspectral approaches for early basal stem rot detection, while Yong et al. (2023) combine hyperspectral imaging and deep learning for automated disease recognition.

These studies illustrate both opportunity and limitation. Physiological models may perform well under a defined sensing protocol, yet deployment requires robustness to differences in cultivar, canopy age, disease severity, illumination, sensor configuration, and environmental background. A foundation model trained on diverse healthy and stressed plantation conditions could potentially provide more transferable disease and nutrient representations than a narrowly trained classifier.

Yield and Production Forecasting

Yield forecasting further demonstrates why plantation intelligence cannot rely on imagery alone. Oil-palm yield reflects interactions among historical weather, soil moisture, management, palm age, biological condition, and lagged reproductive processes.

Khan et al. (2022) evaluate machine-learning yield prediction under varying weather and soil-moisture conditions, illustrating the importance of environmental context. Ang et al. (2022) integrate multiple data sources for block-scale oil-palm yield prediction, while Jamshidi et al. (2024) compare 17 machine-learning and deep-learning models using a comprehensive agronomic dataset.

These studies imply that an effective foundation model should be able to combine unstructured imagery with structured variables. Yield is therefore a demanding downstream test of whether a latent representation captures meaningful plantation state rather than visual correlation alone.

Fruit Maturity and Harvest Intelligence

Fruit-level AI operates at a very different spatial scale from satellite plantation mapping. Suharjito et al. (2021) demonstrate fresh-fruit-bunch ripeness classification on mobile devices, highlighting the possibility of edge deployment. Junior and Suharjito (2023) extend this work to video-based detection, enabling more continuous harvesting support.

Shared data resources are also emerging. Suharjito et al. (2023) publish annotated oil-palm fruit-bunch pile datasets for deep-learning ripeness grading, improving reproducibility and providing reusable training material. Puttinaovarat et al. (2024) connect fruit-bunch classification with geospatial plantation verification,

indicating that fruit-level and estate-level intelligence can increasingly be integrated.
This hierarchical relationship is important for foundation-model design. PALM-FM should not assume that one spatial resolution serves all applications. Instead, it should support representations from regional landscapes through estate blocks and individual palms to fruit-level imagery.

Method

Review Design

This study uses a qualitative literature review rather than a systematic literature review. Its purpose is interpretive integration, theory building, and conceptual architecture development rather than exhaustive retrieval, PRISMA-based screening, bibliometric enumeration, or quantitative meta-analysis.

A qualitative approach is appropriate because the topic spans multiple literatures that have developed independently. Oil-palm remote sensing, disease diagnosis, yield forecasting, agricultural foundation models, domain adaptation, self-supervised learning, multimodal Earth observation, explainable AI, and data governance use different technical terminology and address different units of analysis. The research problem therefore requires synthesis across disciplinary boundaries rather than a narrowly delimited systematic evidence question.

The review focuses primarily on peer-reviewed literature published between 2020 and 2026. Particular emphasis is placed on reputable international journals in remote sensing, agricultural engineering, artificial intelligence, precision agriculture, information fusion, geospatial science, and plantation-specific applications.

Literature Identification

Literature was identified conceptually around eight clusters: foundation models in agriculture; geospatial and remote-sensing foundation models; self-supervised and masked representation learning; multimodal learning; domain adaptation and domain generalization; oil-palm mapping and inventory; physiological, disease, yield, and fruit applications; and responsible AI. These clusters served as sensitizing concepts for comparing how different research communities address representation reuse, transferability, operational deployment, and governance.

The review deliberately incorporates both oil-palm-specific studies and transferable evidence from broader agricultural and Earth-observation research. This is necessary because no single oil-palm literature currently contains all components required to conceptualize a domain-specific foundation model.

References were selected for their conceptual relevance to representation learning, multimodal sensing, temporal modelling, domain transfer, operational plantation tasks, or governance. The intention is not to claim that every publication on these subjects has been retrieved. Instead, the analysis seeks theoretical saturation across the principal components of the proposed PALM-FM architecture.

Thematic Coding and Analytical Synthesis

Each publication was interpreted using initial codes including task, sensing modality, temporal scale, geographic scale, annotation requirement, transfer setting, downstream objective, and reported limitation. These initial codes were subsequently combined into higher-order categories.

Eight themes emerged: task-specific fragmentation; domain generalization; multimodality; plantation lifecycle representation; self-supervised learning; shared downstream adaptation; benchmarking; and governance. These themes structure the Results section.

The Discussion subsequently integrates these findings into PALM-FM. Rather than deriving a statistical summary, the analysis asks how insights from separate literatures can be assembled into a unified technical and institutional architecture.

Results

Theme 1: Oil-Palm AI Has Developed Task Depth without Equivalent Representation Integration

The first major finding is that oil-palm artificial intelligence has achieved considerable specialization. Models now address plantation detection, individual palm recognition, density, canopy characteristics, disease, nutrient status, yield, stand age, height, fruit maturity, and operational verification. However, the architecture of innovation remains strongly task oriented.

This produces duplicated representation learning. A tree detector learns canopy geometry from labelled imagery, while an age-estimation model may independently relearn similar structural patterns. Disease classifiers learn deviations in spectral behaviour without systematically transferring these representations to nutrient or stress models. Yield models reconstruct plantation condition from agronomic variables even when related spatial representations have already been learned by other systems.

The problem is therefore not a lack of models but a lack of reusable representation infrastructure. Foundation-model research in agriculture explicitly seeks to address this problem by separating large-scale representation learning from downstream adaptation (Li et al., 2024; Yin et al., 2025). The strong performance of foundation representations in disease and weed applications suggests that agricultural tasks can benefit from pretraining beyond the immediate target dataset (Espejo-Garcia et al., 2025).

MetaFruit provides an additional indication that agricultural datasets can be organized around broad representation objectives rather than fixed closed-set classification (Li et al., 2025). For oil palm, the equivalent transition would mean that mapping, inventory, disease, yield, and fruit systems no longer begin from independently initialized models.

Theme 2: Generalization Is the Hidden Bottleneck

The literature consistently indicates that high accuracy within the original data domain does not guarantee operational transfer. Random train–validation–test splitting can produce impressive results because observations from the same estate, season, sensor, or experimental protocol appear throughout the dataset. Such evaluation does not necessarily reveal performance on truly unseen operational conditions.

Domain-adaptation research provides extensive evidence for this problem. Peng et al. (2022) show that remote-sensing classification is vulnerable to distribution differences between source and target domains. Wang et al. (2023) demonstrate that agricultural temporal–spectral distributions differ sufficiently across space and time to require explicit unsupervised adaptation. Zhang et al. (2022) similarly apply adversarial domain adaptation to agricultural land extraction, while Rusňák et al. (2023) investigate crop mapping across space and time without target labels.

Source-free adaptation adds another operationally relevant dimension. Mohammadi et al. (2024) show that cross-regional and cross-time crop mapping can be improved without requiring

source-domain data during adaptation. This characteristic is particularly attractive to commercial plantation settings where the original training data may be unavailable because of confidentiality, data residency, or licensing constraints.

Cross-sensor transfer presents an additional challenge. Wang et al. (2022) show that differences between airborne and spaceborne observations can require explicit domain alignment. Oil-palm applications face analogous variation across Sentinel, Landsat, commercial imagery, UAV cameras, smartphones, hyperspectral instruments, and radar platforms.

The finding is therefore that generalization must be treated as a primary performance dimension. A model that achieves slightly lower source-domain accuracy but preserves performance across new estates may be operationally more valuable than a model optimized narrowly for one benchmark.

Theme 3: Tropical Domain Shift Is Multidimensional

Oil-palm domain shift is more complex than a simple transfer between geographical locations. Eight interacting dimensions can be distinguished.

First, **geographic shift** arises across estates, provinces, countries, terrain types, and surrounding landscapes. Abramowitz et al. (2023), for example, demonstrate that distinguishing oil palm from natural forest in Ghana presents a different landscape context from many Southeast Asian plantation studies.

Second, **climatic shift** reflects rainfall, drought, haze, humidity, flooding, temperature, and radiation. Such changes modify canopy moisture, reflectance, physiological condition, and yield.

Third, **seasonal shift** changes illumination, ground conditions, vegetation background, and sensor availability. Persistent cloud also changes which optical observations are available.

Fourth, sensor shift occurs across imaging systems with different resolutions, spectral bands, noise characteristics, and acquisition geometries. Because sensor characteristics alter scale and signal statistics, such mismatch can reduce transferability unless harmonization or domain adaptation is applied (Wang et al., 2022).

Fifth, **plantation lifecycle shift** distinguishes immature, mature, aging, and replanted stands. Palm height, crown size, spacing visibility, and understory exposure change substantially through the lifecycle.

Sixth, soil and terrain shift arises between peat and mineral soils, flat and sloping terrain, and differing drainage environments. These conditions can change background reflectance, moisture regimes, accessibility, and plant-stress expression, thereby altering both sensed features and management responses.

Seventh, management shift results from variation in planting material, density, fertilization, pruning, weed control, road structure, harvesting intensity, and disease management. Such differences can produce distinct canopy structures and operational signals even within otherwise similar climatic zones.

Eighth, **production-system shift** separates industrial estates from smallholder systems. Smallholder resilience research demonstrates that growers operate under heterogeneous technical, financial, institutional, and market constraints, implying that datasets from large estates should not automatically be treated as representative of the entire sector (Hendrawan et al., 2024).

These shifts can occur simultaneously. PALM-FM must therefore be designed for compound domain shift rather than assuming that one adaptation dimension can be isolated from all others.

The transition from fragmented models toward reusable representation learning is illustrated in **Figure 1**. The left side represents the prevailing architecture in which mapping, counting, disease, nutrient, yield, age, and maturity applications are trained separately. The right side represents PALM-FM, in which a common pretrained backbone supports multiple downstream adapters.

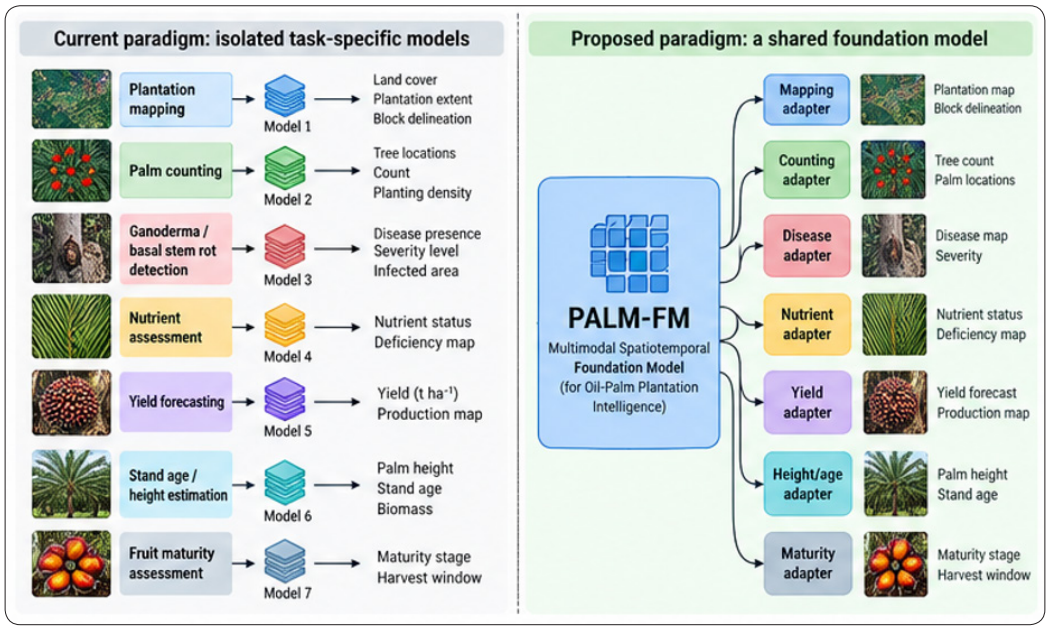


Figure 1: From Task-Specific Oil-Palm AI to Foundation-Model-Based Plantation Intelligence

The figure highlights an important distinction. PALM-FM does not eliminate downstream specialization; rather, it relocates specialization from the complete model to relatively lightweight task-specific components. Improvements to the backbone can therefore potentially benefit several plantation applications simultaneously.

Theme 4: Plantation Intelligence Is Intrinsically Multimodal
Oil-palm systems cannot be represented adequately through a single sensing modality. Optical imagery contributes spectral and textural information, but tropical cloud cover can interrupt observation. Radar contributes all-weather capability and complementary structural or moisture sensitivity. Hyperspectral data provide narrow-band information useful for physiological interpretation. LiDAR and related altimetry characterize vertical structure. UAV imagery provides high spatial resolution, while operational databases contribute information that sensors cannot directly observe.

Oil-palm evidence already demonstrates these complementarities. Mohd Najib et al. (2020) show the benefits of active and passive sensor integration for plantation mapping, while hyperspectral studies demonstrate physiological sensitivity for nutrient and disease assessment (Amirruddin et al., 2020; Azmi et al., 2020; Lee et al., 2022; Yong et al., 2023). UAV imagery supports detailed tree detection at a scale unavailable from medium-resolution satellites (Lee et al., 2024). More broadly, geospatial-foundation-model research combining SAR, optical, and climate information has improved estimation of crop water content and height across heterogeneous fields, providing a close analogue for the multimodal logic proposed for PALM-FM (Hashemi et al., 2025).

Multimodal integration should nevertheless move beyond feature concatenation. Data streams need to be aligned spatially, temporally, and semantically. A rainfall event may influence plant condition weeks or months later. Block-level yield information must be connected to imagery covering the same management unit. Fruit-level photographs should be associated with harvest context rather than treated as unrelated visual samples.

Remote-sensing contrastive learning and cross-modal alignment offer technical pathways for such integration (Berg et al., 2024; Zavras et al., 2025). The broader agricultural image-fusion literature also demonstrates that complementary sensor information can improve crop interpretation when fusion is designed carefully (Li et al., 2021).

Theme 5: Oil Palm Requires a Spatiotemporal Rather than Static Representation
The perennial nature of oil palm means that biological and operational information unfolds across multiple time horizons. A plantation planted twenty years earlier contains historical information that influences current structure and productivity. Weather influences may operate through delayed effects on flowering, bunch formation, and yield. Chronic disease develops progressively rather than instantaneously.

Oil-palm phenology and age studies provide strong evidence for temporal modelling. Effendi et al. (2022) emphasize the opportunities and challenges associated with phenological remote sensing. Zang et al. (2023) link stand age with structural height observations. Weerakitikul et al. (2025) demonstrate that long satellite archives can reconstruct stand age, while Kuapanich et al. (2026) show the importance of temporally updated plantation mapping.

Broader agricultural time-series research supports the same conclusion. Attention-based Sentinel-2 models improve crop mapping by exploiting temporal signatures rather than static scenes (Wang et al., 2021). CropSTS and AgriFM take this principle further by embedding temporal modelling into agricultural foundation-model architecture itself (Yan et al., 2025; Li et al., 2026a).

PALM-FM should therefore represent at least three temporal scales. Short-term representations should capture weather anomalies, stress, harvest operations, and changing observation conditions. Medium-term representations should capture seasonal productivity, fertilization response, and disease progression. Long-term representations should encode establishment, maturity, aging, and replanting.

Theme 6: Self-Supervised Learning Can Reduce the Oil-Palm Annotation Bottleneck
Expert-labelled plantation data are expensive. Disease diagnosis can require pathologists. Nutrient labels may require laboratory analysis. Yield labels require consistent operational recording. Tree inventories and fruit annotations are labour intensive. Large-area mapping requires extensive reference data.

The imbalance between abundant unlabelled data and scarce expert labels makes self-supervised learning particularly attractive. Masked reconstruction allows models to learn from imagery without class labels, while temporal prediction can exploit repeated observations. Cross-modal contrastive objectives can align co-located data streams.

Remote-sensing examples demonstrate the feasibility of these strategies. SpectralMAE learns representations by reconstructing masked hyperspectral information (Zhu et al., 2023). Feature-guided masked autoencoding supports self-supervised Earth-observation representation learning (Wang et al., 2024). Liu et al. (2024) demonstrate self-supervised reconstruction across harmonized satellite time series.

Agricultural foundation models extend these concepts toward operational tasks. Espejo-Garcia et al. (2025) demonstrate that pretrained foundation features can be adapted efficiently using LoRA and distilled into smaller architectures. AgriFM shows how large global remote-sensing corpora can be converted into reusable agricultural representations (Li et al., 2026a).

The practical implication is that oil-palm operators may possess substantially more usable AI training material than conventional supervised pipelines suggest. Historical imagery and operational records should be viewed as pretraining assets even when they were never manually labelled.

Theme 7: A Shared Representation Could Support a Downstream Plantation Ecosystem
The oil-palm literature provides a diverse suite of downstream tasks through which PALM-FM could be evaluated. Plantation mapping and density estimation test large-area spatial representation (Shaharum et al., 2020; Xu et al., 2021; Rodríguez et al., 2021). Individual-tree detection tests high-resolution spatial precision (Wibowo et al., 2022; Lee et al., 2024).

Disease and nutrient assessment test whether the representation captures subtle physiological variation (Amirruddin et al., 2020; Lee et al., 2022; Yong et al., 2023). Yield prediction tests whether the representation can integrate imagery with structured agronomic and environmental variables (Khan et al., 2022; Ang et al., 2022; Jamshidi et al., 2024).

Fruit ripeness tests a smaller spatial scale and offers a path toward edge deployment (Suharjito et al., 2021; Junior and Suharjito, 2023). The availability of annotated fruit datasets improves the possibility of comparing pretrained and task-specific approaches reproducibly (Suharjito et al., 2023). Geospatial verification platforms further illustrate how local observations can be connected to broader plantation intelligence (Puttinaovarat et al., 2024).

The foundation-model hypothesis is therefore empirically testable. PALM-FM should outperform or approach strong task-specific baselines while requiring fewer labelled target samples and preserving greater performance under domain shift.

Theme 8: Benchmarking and Governance Must Be Designed Together

A foundation model cannot be validated solely through conventional random data splitting. Benchmark design must reveal whether representations transfer.

Geographic holdouts should exclude complete estates or regions. Temporal holdouts should exclude complete years or seasons. Sensor holdouts should evaluate previously unseen acquisition systems. Lifecycle holdouts should test young, mature, and aging plantations independently. Production-system holdouts should deliberately include smallholder landscapes.

Domain-generalization performance should be accompanied by uncertainty estimates. A model should ideally become less confident when it encounters conditions outside its experience rather than producing highly certain but incorrect outputs. Explainable AI can help users understand influential observations and features, although explanation methods must be interpreted carefully rather than being treated automatically as causal evidence (Ryo, 2022; Cartolano et al., 2024).

Responsible deployment also requires institutional trust. Agricultural AI can change access to data, advisory power, and decision authority, meaning accountability cannot be reduced to model accuracy (Alexander et al., 2024). Privacy and commercial confidentiality further complicate the creation of large shared datasets.

Federated learning may therefore become important for PALM-FM. Agricultural federated-learning research suggests that distributed model training can enable collaboration without centralizing all raw data, although technical and governance challenges remain (Dembani et al., 2025).

The results answer the first research question by showing that the limitation of current oil-palm AI is increasingly architectural rather than algorithmic. Task-specific modelling has been extremely productive, but each new application tends to reproduce several stages already performed elsewhere: data preparation, feature learning, model optimization, and representation validation.

This duplication becomes costly as the number of applications grows. Plantation mapping, yield forecasting, disease detection, age estimation, fruit maturity, and sustainability monitoring are not biologically independent problems. They describe different manifestations of the same plantation system.

A common representation offers a route to accumulated intelligence. For example, canopy structure learned during tree detection may contribute to stand-age prediction. Temporal anomalies associated with disease could contribute to yield-risk models. Geographic embeddings learned during plantation mapping could provide context for sustainability verification.

The transition should therefore be understood as movement from **model accumulation** toward **representation accumulation**. A foundation model becomes valuable when the knowledge gained from one dataset or task can be retained and reused rather than discarded when the next model is trained.

Proposed PALM-FM Architecture

PALM-FM is conceptualized around six integrated layers. Together, these layers connect heterogeneous plantation observations to reusable representations, task-specific adaptation, and operational safeguards.

The first layer is the **Plantation Multimodal Data Commons**. It includes optical satellite imagery, SAR imagery, UAV RGB and multispectral imagery, hyperspectral observations, LiDAR or canopy-height measurements, weather sequences, soil and terrain characteristics, planting dates, fertilizer records, harvest records, disease observations, and other operational information.

The second layer contains **modality-specific encoders**. Separate encoders are appropriate because radar backscatter, hyperspectral signatures, imagery, structured weather sequences, and textual agronomic information follow different statistical structures.

The third layer performs **spatiotemporal and cross-modal alignment**. Observations should be linked according to plantation block, palm coordinates, observation date, lifecycle stage, and management history. Cross-attention or contrastive objectives can be used to associate modalities that describe the same underlying plantation condition.

The fourth layer conducts **large-scale self-supervised pretraining**. Relevant objectives include masked image reconstruction, masked temporal reconstruction, cross-modal contrastive learning, future-embedding prediction, temporal-order prediction, and teacher–student distillation.

The fifth layer is a **shared plantation embedding** representing common information about plantation condition. This embedding forms the core PALM-FM backbone.

The sixth layer contains **lightweight downstream adapters** for specific tasks. Mapping, tree counting, disease detection, nutrient status, yield prediction, age, height, fruit maturity, replanting, and sustainability applications can each use separate adaptation heads while retaining the common backbone.

The complete architecture is summarized in Figure 2, which illustrates how heterogeneous plantation observations are encoded, aligned, pretrained, and converted into a reusable representation before being adapted to downstream tasks. The diagram also makes explicit that shared representation learning occurs before lightweight specialization for mapping, counting, disease, nutrition, yield, structure, and fruit-maturity applications.

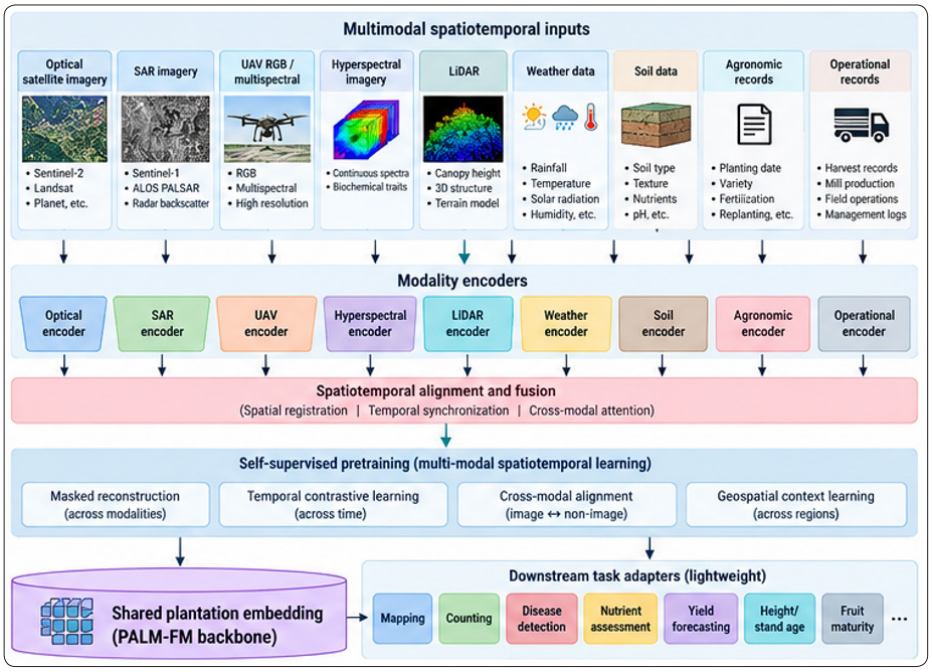


Figure 2: PALM-FM Multimodal Spatiotemporal Architecture

The architecture is deliberately modular. A plantation without LiDAR should still be able to use the model with optical, radar, weather, and operational information. Missing-modality robustness should therefore be included during pretraining rather than treated as an exceptional deployment failure.

Tropical Domain Shift as the Defining Scientific Problem

The second research question asks which domain shifts are most important. The evidence suggests that no single shift dominates universally because the shifts interact. Geographic transfer often contains simultaneous differences in climate, age distribution, soil, management, and sensor availability.

For this reason, tropical domain shift should be represented as a multidimensional state rather than a binary source–target label. A target estate can be similar to the source in one dimension and radically different in another. Two estates may use the same satellite sensor but differ in palm age and rainfall. Conversely, neighbouring estates may share climate and soils while using different UAV systems and management practices.

This distinction has consequences for model evaluation. If a model fails in an unseen region, researchers should attempt to identify whether the primary cause is spectral variation, phenological displacement, sensor mismatch, management difference, or compound shift. Such diagnosis can guide the choice among domain adaptation, additional pretraining, sensor harmonization, or collection of targeted labels.

Figure 3 conceptualizes tropical domain shift as a matrix connecting environmental and operational sources of variation to changes in observed data distributions and subsequent model-performance risks. It therefore provides a compact diagnostic framework for interpreting why an apparently strong source-domain model may deteriorate after field transfer.

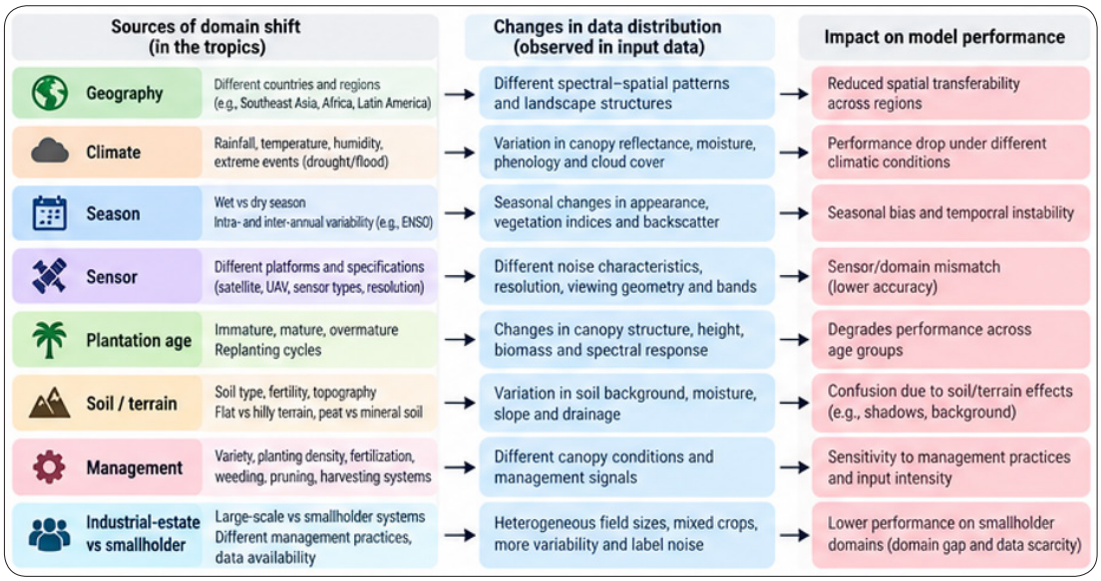


Figure 3: Tropical Domain Shift Matrix for Oil-Palm Artificial Intelligence

The matrix provides a practical reporting framework. Future oil-palm AI studies should identify which dimensions are represented during training, which are held out during validation, and which remain unknown. This would substantially improve the interpretability of generalization claims.

Designing a Meaningful PALM-FM Benchmark

The sixth research question concerns evaluation. A PALM-FM benchmark should contain at least five levels.

The first is **in-domain performance**, retaining conventional random or spatially balanced test sets for comparison with existing methods. This establishes whether the representation is technically competitive.

The second is **cross-domain generalization**. Complete estates, regions, years, seasons, sensors, or plantation age classes should be withheld from pretraining or task adaptation as appropriate.

The third is **label-efficiency evaluation**. Downstream models should be trained with progressively smaller labelled fractions, such as 100%, 10%, 5%, 1%, and few-shot examples. A foundation model should provide increasing relative advantage as labels become scarce.

The fourth is **missing-modality and degradation testing**. Optical imagery might be unavailable because of clouds, an operational dataset may contain gaps, or hyperspectral information may not exist at the target estate. The model should degrade gracefully rather than fail catastrophically.

The fifth is **confidence and uncertainty evaluation**. Standard accuracy should be complemented by calibration, out-of-distribution detection, confidence–error relationships, and uncertainty under domain shift.

Different downstream tasks require different metrics. Segmentation can use intersection-over-union and F1-score. Tree detection can use average precision and counting error. Disease and maturity classification can emphasize macro-F1 and sensitivity. Yield, age, or height prediction can use mean absolute error, root mean square error, and coefficient of determination.

A particularly useful metric would be the **domain-generalization gap**, defined as the loss in performance between familiar and genuinely unseen domains. This shifts attention away from the highest single accuracy score toward operational robustness.

Multimodal Pretraining versus Conventional Sensor Fusion

The third and fourth research questions concern multimodal architecture and learning mechanisms. Conventional sensor fusion generally begins with a predefined downstream target. Features from optical, radar, or other sources are combined specifically to improve that target.

PALM-FM proposes a different objective. Cross-modal relationships should be learned **before** the downstream task is known. An optical image and radar observation from the same block and time period should be recognized as related descriptions of the same plantation state. A rainfall sequence and later canopy response can provide temporal alignment signals. A disease observation can be associated with preceding remote-sensing trajectories.

Contrastive learning is especially suitable for such alignment. Positive pairs can represent the same location observed by different sensors, whereas negative pairs represent different plantation states. Masked cross-modal prediction provides another mechanism: the model could infer missing optical features from radar and temporal context or reconstruct environmental variables from shared embeddings.

This approach also creates the possibility of multimodal substitution. If the model learns stable relationships among sensors during pretraining, it may remain useful when one modality is missing at deployment. Such robustness is essential if PALM-FM is intended to serve both highly instrumented estates and data-scarce smallholder environments.

Spatiotemporal Memory and Plantation Lifecycle Intelligence

The fourth research question also requires considering temporal architecture explicitly. Oil palm is not simply a repeatedly observed object; it is a changing biological asset with memory.

A useful PALM-FM representation should encode the history of a block. The same spectral anomaly can have different meanings depending on whether a stand is newly planted, mature, recovering from drought, recently fertilized, or experiencing chronic disease.

Temporal attention, recurrent architectures, state-space models, or video-style transformers could potentially provide this memory. AgriFM's synchronized spatiotemporal architecture demonstrates that agricultural foundation modelling benefits from processing spatial and temporal characteristics jointly rather than as independent branches (Li et al., 2026a). CropSTS similarly emphasizes temporal structure as a fundamental design element (Yan et al., 2025).

At a conceptual level, PALM-FM should therefore learn a **plantation trajectory embedding**. Such an embedding would represent not merely what a block looks like today but how it has arrived at its current state.

This creates possibilities for early warning. A block whose trajectory diverges from comparable healthy blocks could be flagged for field inspection even when the exact stress has not yet been classified. Such anomaly detection could become an important application of foundation representations.

Parameter-Efficient Adaptation and Field Deployment

A foundation model does not imply that every field device must execute the complete pretrained architecture. Large-scale pretraining can occur centrally, while task-specific deployment uses lightweight adaptations.

LoRA and related parameter-efficient methods reduce the number of parameters that must be updated for a downstream domain. Knowledge distillation can subsequently compress foundation representations into smaller networks suitable for mobile, UAV, or edge deployment, as demonstrated in agricultural vision research (Espejo-Garcia et al., 2025).

This architecture is especially relevant to the oil-palm sector because connectivity and computing resources vary substantially. A central plantation analytics centre may operate powerful GPU infrastructure, whereas a field harvester may use only a smartphone.

PALM-FM should therefore be designed as an ecosystem rather than a single monolithic application. The ecosystem could include a high-capacity pretraining model, estate-scale inference services, compressed UAV models, and small mobile adapters.

Governance, Privacy, and Responsible AI

The seventh research question concerns institutional requirements. Technical development will not succeed if governance is postponed until after model construction.

Oil-palm data are commercially and politically sensitive. Yield records may reveal productivity. Disease records can have reputational implications. High-resolution imagery may expose operational infrastructure. Smallholder data can create concerns about surveillance, market access, or unequal bargaining power.

Agricultural responsible-AI literature emphasizes that responsibility must be allocated across technology developers, data owners, users, and institutions rather than attributed vaguely to "the algorithm" (Alexander et al., 2024). Explainability can increase transparency but should be designed around the needs of agronomists and field managers rather than only machine-learning developers (Ryo, 2022; Cartolano et al., 2024).

Federated learning offers one pathway for cross-company or cross-institutional collaboration. Instead of pooling all raw data centrally, participating institutions can train locally and exchange model updates. However, federated learning does not automatically solve governance. Questions remain regarding participation, aggregation rights, model ownership, benefit sharing, security, and representation (Dembani et al., 2025).

Smallholders require particular consideration. If PALM-FM is pretrained mainly using highly organized corporate estates, it may perform poorly in irregular smallholder landscapes. This would create a representational bias precisely where digital infrastructure is weakest.

Responsible PALM-FM development should therefore include smallholder landscapes intentionally, document geographic and management composition, report subgroup performance, and establish mechanisms through which participating producers benefit from the resulting intelligence. These requirements should be incorporated into dataset documentation and model-governance protocols rather than treated as optional post-deployment safeguards.

From Prediction toward Integrated Plantation Intelligence

The broader significance of PALM-FM lies in moving beyond a collection of predictions. If a common representation captures multimodal plantation state, it can become the perception layer of a larger decision-support environment.

For example, a block-level representation could combine palm age, current canopy condition, recent rainfall, historic yield, disease observations, and management records. A yield adapter could estimate production, while a disease adapter assesses risk and a replanting adapter evaluates lifecycle status. Because these outputs originate from a shared representation, contradictory predictions can be detected and investigated.

A foundation representation could also contribute to digital twins. A plantation digital twin requires an updated state representation linking physical observations with computational models, and agricultural digital-twin research similarly emphasizes the integration of continuous sensing, virtual representations, and model-based decision support (Purcell and Neubauer, 2023). PALM-FM could provide the observational embedding through which remote sensing, field measurements, and operational records are continuously assimilated.

Another future direction is vision–language interaction. Plantation managers could query blocks using natural language, ask why an anomaly was detected, retrieve similar historical cases, or connect visual evidence with agronomic documentation. The development of remote-sensing vision–language models suggests that such interaction is technically plausible, although strong domain grounding will be required to prevent unsupported recommendations (Zavras et al., 2025).

Research and Implementation Roadmap

A realistic development strategy should proceed incrementally. This sequencing reduces technical and governance risk by requiring data interoperability and cross-domain validation before progressively more complex operational deployment.

Phase 1: Data Commons and Interoperability. Participating institutions should establish shared metadata standards for location, date, sensor, planting age, management unit, label provenance, and quality. Existing datasets should be catalogued and licensing conditions clarified.

Phase 2: Self-Supervised Pretraining. Large collections of optical, radar, UAV, hyperspectral, and operational data should be used for unimodal and initial multimodal pretraining. Representation quality should be evaluated across several downstream tasks.

Phase 3: Multimodal and Temporal Alignment. Co-located and temporally matched observations should be used to learn cross-modal relationships. The model should be trained intentionally with missing modalities so that incomplete target environments can be supported.

Phase 4: Cross-Domain Benchmarking. Estates, regions, years, sensors, palm ages, and management types should be held out systematically. Standardized protocols should make domain-generalization claims directly comparable.

Phase 5: Multi-Task Operational Deployment. Mature downstream adapters should be integrated into plantation decision-support systems. Human validation and uncertainty thresholds should govern high-consequence actions.

Phase 6: Continual and Federated Learning. The model should be updated as new regions, sensors, planting materials, diseases, and climatic conditions emerge. Continual learning should minimize catastrophic forgetting while federated mechanisms enable collaboration where raw data cannot be centralized.

The proposed development pathway is summarized in Figure 4. The roadmap emphasizes that an oil-palm foundation model is not created merely by scaling the number of neural-network parameters; it requires coordinated development of data infrastructure, representation learning, benchmarking, deployment, and governance.

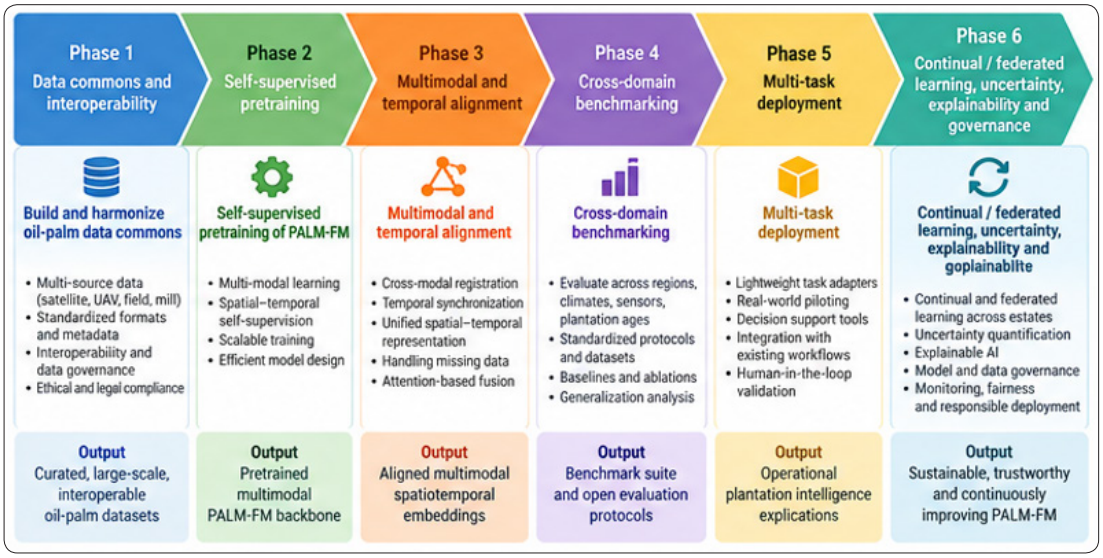


Figure 4: Research and Deployment Roadmap for PALM-FM

The final phase should not be considered an endpoint. Tropical plantation environments continue to change, sensors are replaced, management practices evolve, and climate conditions shift. PALM-FM must therefore be treated as continually maintained digital infrastructure.

Implications for Research, Industry, and Public Policy

For researchers, the central implication is that external validation should become more prominent. Future studies should report not only random-split accuracy but performance across estates, years, sensors, and management systems. Dataset descriptions should state geographic and temporal coverage clearly enough for readers to evaluate generalization claims.

For plantation companies, historical data should be treated as strategic AI infrastructure. Decades of yield, weather, fertilizer, disease, imagery, and operational records may provide enormous pretraining value even if they were not originally collected for machine learning. Consistent metadata, georeferencing, and data governance therefore have long-term economic value.

For technology developers, the opportunity lies in modularity. Instead of selling isolated algorithms, developers could build adapters and applications around a validated plantation foundation model. Such modularity could lower entry barriers for new use cases.

For governments and industry associations, a major contribution would be the development of neutral benchmark datasets. Geographically diverse validation resources could improve transparency, reduce duplicated labelling, and ensure that smallholder and non-industrial landscapes are represented.

The broader policy implication is that foundation-model infrastructure should not become an additional source of digital concentration. Public research institutions, universities, producer organizations, and smallholder representatives should participate in governance and benchmark development.

Conclusion and Recommendations

Oil-palm artificial intelligence is entering a stage in which continued progress will depend not only on creating more accurate task-specific algorithms but on building representations that can be reused across tasks, locations, sensors, and time. Existing research already demonstrates substantial capability in plantation mapping, tree inventory, disease detection, nutrient assessment, yield forecasting, age and height estimation, and fruit maturity evaluation. The remaining challenge is to connect these capabilities into an integrated and transferable intelligence architecture.

This article proposes PALM-FM as a multimodal spatiotemporal foundation model for oil-palm plantation intelligence. The framework treats satellite imagery, radar, UAV observations, hyperspectral information, structural sensing, weather, soil, agronomic records, production histories, and management information as complementary descriptions of a shared plantation state. Self-supervised pretraining can transform abundant unlabelled data into reusable representations, while downstream adapters can specialize the shared backbone for operational tasks.

Tropical domain shift should be treated as the central scientific test of this architecture. Geographic variation, climatic conditions, seasonality, sensor differences, plantation lifecycle, soils, management practices, and contrasts between industrial estates and smallholder systems can all change the data distribution encountered by artificial intelligence. Consequently, cross-estate, cross-region, cross-year, cross-sensor, and low-label evaluation should become standard components of future oil-palm AI research.

A successful PALM-FM ecosystem will require more than model development. Interoperable data standards, uncertainty estimation, explainability, privacy protection, federated collaboration, human oversight, representative smallholder data, and transparent governance must form part of the architecture from the beginning. Foundation-model development should therefore be understood as a socio-technical infrastructure project.

Future research should prioritize geographically diverse data commons, multimodal and temporal self-supervision, explicit domain-generalization benchmarks, parameter-efficient adaptation, edge deployment, continual learning, federated pretraining, and links with plantation digital twins. A particularly important frontier will be the integration of vision, sensor, operational, and agronomic knowledge so that plantation intelligence progresses from isolated predictions toward contextual decision support.

The strategic transition proposed here is therefore from **task-specific oil-palm AI to reusable plantation intelligence**. If developed responsibly, PALM-FM could provide a common foundation for scientific research, precision management, sustainability monitoring, smallholder support, autonomous

inspection, and future intelligent plantation systems.

References

1. Abade A, Ferreira PA, Vidal FB (2021) Plant diseases recognition on images using convolutional neural networks: A systematic review. Computers and Electronics in Agriculture 185: 106125.
2. Abramowitz J, Cherrington E, Griffin R, Muench R, Mensah F (2023) Differentiating oil palm plantations from natural forest to improve land cover mapping in Ghana. Remote Sensing Applications: Society and Environment 30: 100968.
3. Alexander CS, Yarborough M, Smith A (2024) Who is responsible for "responsible AI"?: Navigating challenges to build trust in AI agriculture and food system technology. Precision Agriculture 25: 146–185.
4. Amirruddin AD, Muharam FM, Ismail MH, Tan NP, Ismail MF (2020) Hyperspectral spectroscopy and imbalance data approaches for classification of oil palm's macronutrients observed from frond 9 and 17. Computers and Electronics in Agriculture 178: 105768.
5. An X, He W, Zou J, Yang G, Zhang H (2024) Pretrain a remote sensing foundation model by promoting intra-instance similarity. IEEE Transactions on Geoscience and Remote Sensing 62: 1–15.
6. Ang Y, Shafri HZM, Lee YP, Bakar SA, Abidin H, Mohd Junaidi MUU, Hashim SJ, Che'Ya NN, Hassan MR, Lim HS, Abdullah R, Yusup Y, Muhammad SA, Teh SY, Samad MN (2022) Oil palm yield prediction across blocks from multi-source data using machine learning and deep learning. Earth Science Informatics 15: 2349–2367.
7. Azmi ANN, Bejo SK, Jahari M, Muharam FM (2020) Early detection of Ganoderma boninense in oil palm seedlings using support vector machines. Remote Sensing 12(23): 3920.
8. Berg P, Uzun B, Pham MT, Courty N (2024) Multimodal supervised contrastive learning in remote sensing downstream tasks. IEEE Geoscience and Remote Sensing Letters 21: 6007405.
9. Cartolano A, Cuzzocrea A, Pilato G (2024) Analyzing and assessing explainable AI models for smart agriculture environments. Multimedia Tools and Applications 83: 37225–37246.
10. da Silveira F, Lermen FH, Amaral FG (2021) An overview of agriculture 4.0 development: Systematic review of descriptions, technologies, barriers, advantages, and disadvantages. Computers and Electronics in Agriculture 189: 106405.
11. Dembani R, Karvelas I, Akbar NA, Rizou S, Tegolo D, Fountas S (2025) Agricultural data privacy and federated learning: A review of challenges and opportunities. Computers and Electronics in Agriculture 232: 110048.
12. Effendi RH, Wikantika K, Darmawan S (2022) Modeling of oil palm phenology based on remote sensing data: Opportunities and challenges. Journal of Applied Remote Sensing 16: 021501.
13. Espejo-Garcia B, Güldenring R, Nalpantidis L, Fountas S (2025) Foundation vision models in agriculture: DINOv2, LoRA and knowledge distillation for disease and weed identification. Computers and Electronics in Agriculture 239: 110900.
14. Hashemi MGZ, Alemohammad H, Jalilvand E, Tan PN, Judge J, Cosh M, Das NN (2025) Estimating crop biophysical parameters from satellite-based SAR and optical observations using self-supervised learning with geospatial foundation models. Remote Sensing of Environment 327: 114825.
15. Hendrawan D, Chrisendo D, Musshoff O (2024) Strengthening oil palm smallholder farmers' resilience to future industrial challenges. Scientific Reports 14: 12105.
16. Hong D, Zhang B, Li X, Li Y, Li C, Yao J, Yokoya N, Li H, Ghamisi P, Jia X, Plaza A, Gamba P, Benediktsson JA, Chanussot J (2024) SpectralGPT: Spectral remote sensing foundation model. IEEE Transactions on Pattern Analysis and Machine Intelligence 46(8): 5227–5244.
17. Huo C, Chen K, Zhang S, Wang Z, Yan H, Shen J, Hong Y, Qi G, Fang H, Wang Z (2025) When remote sensing meets foundation model: A survey and beyond. Remote Sensing 17(2): 179.
18. Jamshidi EJ, Yusup Y, Hooy CW, Kamaruddin MA, Mat Hassan H, Muhammad SA, Shafri HZM, Then KH, Norizan MS, Tan CC (2024) Predicting oil palm yield using a comprehensive agronomy dataset and 17 machine learning and deep learning models. Ecological Informatics 81: 102595.
19. Junior FA, Suharjito (2023) Video based oil palm ripeness detection model using deep learning. Heliyon 9(1): e13036.
20. Khan N, Kamaruddin MA, Sheikh UU, Zawawi MH, Yusup Y, Bakht MP, Mohamed Noor N (2022) Prediction of oil palm yield using machine learning in the perspective of fluctuating weather and soil moisture conditions: Evaluation of a generic workflow. Plants 11(13): 1697.
21. Kok ZH, Shariff ARM, Alfatni MSM, Bejo SK (2021) Support vector machine in precision agriculture: A review. Computers and Electronics in Agriculture 191: 106546.
22. Kuapanich N, Zhang Z, Shi B, Liu J, Jiang J, Huang J, Tan S, Zheng J (2026) Toward 10 m regional-scale time-series oil palm mapping in Malaysia and Indonesia (2020–2024) using Sentinel-2 and noise-robust deep learning from low-resolution historical maps. Forests 17(7): 823.
23. Kwak GH, Park NW (2022) Unsupervised domain adaptation with adversarial self-training for crop classification using remote sensing images. Remote Sensing 14(18): 4639.
24. Lee CC, Koo VC, Lim TS, Lee YP, Abidin H (2022) A multi-layer perceptron-based approach for early detection of BSR disease in oil palm trees using hyperspectral images. Heliyon 8(4): e09252.
25. Lee SS, Lim LG, Palaiahnakote S, Cheong JX, Lock SSM, Ayub MNB (2024) Oil palm tree detection in UAV imagery using an enhanced RetinaNet. Computers and Electronics in Agriculture 227: 109530.
26. Li D, Song Z, Quan C, Xu X, Liu C (2021) Recent advances in image fusion technology in agriculture. Computers and Electronics in Agriculture 191: 106491.
27. Li J, Lammers K, Yin X, Yin X, He L, Sheng J, Lu R, Li Z (2025) MetaFruit meets foundation models: Leveraging a comprehensive multi-fruit dataset for advancing agricultural foundation models. Computers and Electronics in Agriculture 231: 109908.
28. Li J, Xu M, Xiang L, Chen D, Zhuang W, Yin X, Li Z (2024) Foundation models in smart agriculture: Basics, opportunities, and challenges. Computers and Electronics in Agriculture 222: 109032.
29. Li W, Liang S, Chen K, Chen Y, Ma H, Xu J, Ma Y, Zhang Y, Guan S, Fang H, Shi Z (2026a) AgriFM: A multi-source temporal remote sensing foundation model for agriculture mapping. Remote Sensing of Environment 334: 115234.
30. Li X, Li C, Vivone G, Hong D (2026b) SeaMo: A season-aware multimodal foundation model for remote sensing. Information Fusion 125: 103334.
31. Liu H, Zhang HK, Huang B, Yan L, Tran KK, Qiu Y, Zhang X, Roy DP (2024) Reconstruction of seamless harmonized Landsat Sentinel-2 time series via self-supervised learning. Remote Sensing of Environment 308: 114191.
32. Meng S, Wang X, Hu X, Luo C, Zhong Y (2021) Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery. Computers and Electronics in Agriculture 186: 106188.
33. Mohammadi S, Belgiu M, Stein A (2024) A source-free unsupervised domain adaptation method for cross-regional and cross-time crop mapping from satellite image time series. Remote Sensing of Environment 314: 114385.
34. Mohd Najib NE, Kanniah KD, Cracknell AP, Yu L (2020) Synergy of active and passive remote sensing data for effective mapping of oil palm plantation in Malaysia. Forests 11(8): 858.
35. Nyborg JC, Pelletier C, Lefèvre S, Assent I (2022) TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. ISPRS Journal of Photogrammetry and Remote Sensing 188: 301–313.
36. Peng J, Huang Y, Sun W, Chen N, Ning Y, Du Q (2022) Domain

adaptation in remote sensing image classification: A survey. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15: 9842–9859.
37. Purcell W, Neubauer T (2023) Digital twins in agriculture: A state-of-the-art review. Smart Agricultural Technology 3: 100094.
38. Puttinaovarat S, Chai-Arayalert S, Saetang W (2024) Oil palm bunch ripeness classification and plantation verification platform: Leveraging deep learning and geospatial analysis and visualization. ISPRS International Journal of Geo-Information 13(5): 158.
39. Rodríguez AC, D'Aronco S, Schindler K, Wegner JD (2021) Mapping oil palm density at country scale: An active learning approach. Remote Sensing of Environment 261: 112479.
40. Rusňák T, Kasanický T, Malík P, Mojžiš J, Zelenka J, Sviček M, Abrahám D, Halabuk A (2023) Crop mapping without labels: Investigating temporal and spatial transferability of crop classification models using a 5-year Sentinel-2 series and machine learning. Remote Sensing 15(13): 3414.
41. Ryo M (2022) Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. Artificial Intelligence in Agriculture 6: 257–265.
42. Shaharum NSN, Shafri HZM, Ghani WAWAK, Samsatli S, Al-Habshi MMA, Yusuf B (2020) Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. Remote Sensing Applications: Society and Environment 17: 100287.
43. Suharjito, Elwirehardja GN, Prayoga JS (2021) Oil palm fresh fruit bunch ripeness classification on mobile devices using deep learning approaches. Computers and Electronics in Agriculture 188: 106359.
44. Suharjito, Junior FA, Koeswandy YP, Debi, Nurhayati PW, Asrol M, Marimin (2023) Annotated datasets of oil palm fruit bunch piles for ripeness grading using deep learning. Scientific Data 10: 72.
45. Wang J, Ma A, Zhong Y, Zheng Z, Zhang L (2022) Cross-sensor domain adaptation for high spatial resolution urban land-cover mapping: From airborne to spaceborne imagery. Remote Sensing of Environment 277: 113058.
46. Wang Y, Feng L, Zhang Z, Tian F (2023) An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery. ISPRS Journal of Photogrammetry and Remote Sensing 199: 102–117.
47. Wang Y, Hernández Hernández H, Albrecht CM, Zhu XX (2024) Feature guided masked autoencoder for self-supervised learning in remote sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.
48. Wang Y, Zhang Z, Feng L, Ma Y, Du Q (2021) A new attention-based CNN approach for crop mapping using time series Sentinel-2 images. Computers and Electronics in Agriculture 184: 106090.
49. Weerakitikul B, Koedsin W, Ritchie RJ, Kokkaew E, Chan JCW (2025) Multi-sensor remote sensing approach for oil palm mapping and stand age detection using 38-year Landsat and Sentinel time series data in the Google Earth Engine. Geomatica 77(2): 100070.
50. Wibowo H, Sitanggang IS, Mushthofa, Adrianto HA (2022) Large-scale oil palm trees detection from high-resolution remote sensing images using deep learning. Big Data and Cognitive Computing 6(3): 89.
51. Xiao A, Xuan W, Wang J, Huang J, Tao D, Lu S, Yokoya N (2025) Foundation models for remote sensing and Earth observation: A survey. IEEE Geoscience and Remote Sensing Magazine 13(4): 297–324.
52. Xu K, Qian J, Hu Z, Duan Z, Chen C, Liu J, Sun J, Wei S, Xing X (2021) A new machine learning approach in detecting the oil palm plantations using remote sensing data. Remote Sensing 13(2): 236.
53. Yan J, Gu X, Chen Y (2025) CropSTS: A remote sensing foundation model for cropland classification with decoupled spatiotemporal attention. Remote Sensing 17(14): 2481.
54. Yin S, Xi Y, Zhang X, Sun C, Mao Q (2025) Foundation models in agriculture: A comprehensive review. Agriculture 15(8): 847.
55. Yong LZ, Khairunniza-Bejo S, Jahari M, Muharam FM (2023) Automatic disease detection of basal stem rot using deep learning and hyperspectral imaging. Agriculture 13(1): 69.
56. Zang J, Ni W, Zhang Y (2023) Spatially-explicit mapping annual oil palm heights in Peninsular Malaysia combining ICESat-2 and stand age data. Remote Sensing of Environment 295: 113693.
57. Zavras A, Michail D, Demir B, Papoutsis I (2025) Mind the modality gap: Towards a remote sensing vision-language model via cross-modal alignment. ISPRS Journal of Photogrammetry and Remote Sensing 228: 270–287.
58. Zhang J, Xu S, Sun J, Ou D, Wu X, Wang M (2022) Unsupervised adversarial domain adaptation for agricultural land extraction of remote sensing images. Remote Sensing 14(24): 6298.
59. Zhang M, Yang B, Hu X, Gong J, Zhang Z (2024) Foundation model for generalist remote sensing intelligence: Potentials and prospects. Science Bulletin 69(23): 3652–3656.
60. Zhou G, Qian L, Gamba P (2025) Advances on multimodal remote sensing foundation models for Earth observation downstream tasks: A survey. Remote Sensing 17(21): 3532.
61. Zhu L, Wu J, Biao W (2023) SpectralMAE: Spectral masked autoencoder for hyperspectral remote sensing image reconstruction. Sensors 23(7): 3728.

Copyright: © 2026 Loso Judijanto. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.