USOD PUBLISHERS
United Scientific Open Discovery

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

Research Article

Palm FM: Toward an Oil-Palm Foundation Model for Multimodal Spatiotemporal Plantation Intelligence under Tropical Domain ShiftIntegrated Oil Palm Plantation-to-Mill Decision Intelligence

Loso Judijanto

IPOSS Jakarta, Indonesia

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

Citation: Loso Judijanto (2026) Palm FM: Toward an Oil-Palm Foundation Model for Multimodal Spatiotemporal Plantation Intelligence under Tropical Domain Shift. J of Adv Art Int Sys & Frontiers 1(2):1-12.

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

Abstract

Artificial intelligence in oil-palm plantations has developed rapidly, but its dominant paradigm remains fragmented. Separate models are typically developed for palm detection, disease surveillance, nutrient assessment, fruit grading, yield forecasting, plantation mapping and sustainability monitoring, creating repeated annotation requirements and limited transfer across estates, sensors and regions. This qualitative literature review examines whether the next stage of plantation intelligence should instead be organised around an oil-palm-specific multimodal foundation model, termed **Palm FM**. Drawing on peer-reviewed literature published from 2020 to 2026 concerning remote-sensing foundation models, agricultural artificial intelligence, multimodal learning, domain adaptation and oil-palm sensing, the study develops a thematic synthesis rather than a systematic literature review. Seven interconnected themes emerge: task fragmentation, multimodal representation, temporal learning, tropical domain shift, downstream adaptation, generalisation-oriented benchmarking, and trustworthy data governance. The review proposes a five-layer PalmFM architecture comprising multimodal sensing, self-supervised pretraining, plantation representation learning, task-specific adaptation and operational feedback. It further argues that geographic, temporal, sensor, biological and management domain shifts should become explicit benchmarking dimensions rather than nuisance variation hidden by random train-test splits. PalmFM should therefore be evaluated through cross-estate, cross-country, future-year and cross-sensor transfer, few-shot performance, calibration and operational utility. The article concludes that the strategic value of an oil-palm foundation model lies not primarily in model scale but in its ability to create reusable, temporally grounded and governance-compatible plantation representations across heterogeneous tropical production systems.

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

Introduction

Artificial intelligence is increasingly embedded within agricultural observation and decision-support systems. Computer vision, remote sensing, machine learning and deep learning now support crop mapping, stress diagnosis, yield forecasting, phenotyping and increasingly automated management. Nevertheless, much agricultural AI remains organised around individually supervised tasks: a dataset is assembled for one prediction target, a model is trained, and the resulting representation is rarely reused beyond that problem. Foundation models challenge this architecture by learning broadly reusable representations during large-scale pretraining and subsequently adapting those representations to many downstream applications with considerably smaller labelled datasets (Li et al., 2024b; Wang et al., 2023a; Wang et al., 2023b; Cha et al., 2024; Huo et al., 2025; Yin et al., 2025).

This paradigm is particularly relevant to Earth observation. Remote-sensing models must interpret information acquired across radically different physical sensing processes, geographic regions, seasons, spatial resolutions and temporal frequencies. Remote-sensing foundation models have consequently moved beyond conventional natural-image transfer learning toward domain-native pretraining, self-supervised learning, multimodal alignment and task-efficient adaptation. Recent developments encompass remote-sensing-specific vision transformers, vision-language representation learning, billion-parameter scaling, multi-task pretraining, multimodal Earth-observation architectures, SAR-specific foundation models and parameter-efficient multimodal adaptation (Liu et al., 2024a; Wang et al., 2024; Wu et al., 2025; Du et al., 2025; Zhang et al., 2025; Li et al., 2026a).

Agriculture adds a further dimension: time. Crop appearance is not simply a static class but the observable expression of biological development, management intervention and weather history. A useful agricultural foundation model must therefore learn not only spatial structure but seasonality, phenology and cross-sensor temporal correspondence. Recent agricultural foundation-model research demonstrates the importance of large multi-source time series, while season-aware remote-sensing models show that explicitly learning temporal variation can improve transferable geospatial representation (Li et al., 2026b; Li et al., 2026c; Yan et al., 2025). This progression strongly suggests that specialised agricultural sectors may eventually benefit from domain foundation models whose pretraining objectives reflect crop biology and operational information rather than generic scene classification.

Oil palm represents an unusually compelling case. Modern plantations generate an expanding digital footprint from Sentinel-1 SAR, Sentinel-2 optical imagery, commercial very-high-resolution satellites, UAV photography, multispectral and hyperspectral

systems, weather networks, soils, topography, machinery, yield records and field inspections. Yet these information streams remain analytically fragmented. A recent sector-level review similarly found that digital oil-palm applications remain concentrated in mapping, disease detection and fruit grading, underscoring the prevalence of task-specific technology pathways (Bin Mohamad Zaki et al., 2025). Contemporary studies separately address tree detection, counting, growth observation, nutrient status, basal stem rot, fruit maturity, yield, age structure, height and plantation boundaries (Zheng et al., 2020; Zheng et al., 2021; Amirruddin et al., 2020; Kurihara et al., 2022; Chungcharoen et al., 2022; Ang et al., 2022; Zang et al., 2023; Lee et al., 2024). Valuable information learned by one model is therefore rarely preserved as reusable knowledge for another.

Fragmentation is not merely computational inefficiency. It creates a generalisation problem. A classifier trained on one estate may encounter different palm ages, genotypes, soils, fertilisation histories, understorey conditions, disease pressures and harvesting regimes elsewhere. UAV cameras differ in optics and acquisition geometry; commercial satellite platforms differ in spatial and spectral characteristics; Sentinel optical imagery is frequently compromised by tropical cloud; and SAR produces physically distinct observations affected by scattering characteristics. Even within one estate, annual climate variation and canopy maturation modify the data distribution. Cross-regional oil-palm experiments have already shown that domain adaptation and domain generalisation are necessary when models are transferred between geographic and sensing contexts (Zheng et al., 2020; Zheng et al., 2022).

The issue should therefore be reframed. Instead of asking how to build increasingly accurate models for isolated plantation tasks, research should ask what representations an AI system must learn to understand oil-palm production environments across tasks and domains. A foundation model offers a conceptual route to this objective because pretraining can absorb information from large quantities of unlabeled or weakly labelled data and subsequent adaptation can reuse those representations across heterogeneous downstream problems. Self-supervised approaches, masked modelling, contrastive learning, few-shot adaptation and transfer learning provide technical mechanisms for reducing the dependence on costly field annotation (Muhtar et al., 2023; Bi et al., 2023; Alosaimi et al., 2023; Liu et al., 2024b; Ma et al., 2024).

This review therefore develops the concept of **Palm FM**, an oil-palm-specific multimodal and spatiotemporal foundation model. PalmFM is not proposed as one fixed neural architecture. It is instead conceptualised as a research programme for reusable plantation representation learning. The review asks what data PalmFM should learn from, how heterogeneous information should be aligned, how tropical domain shift should be represented, which tasks should be used to evaluate transfer and which institutional arrangements could enable learning across companies without requiring uncontrolled sharing of proprietary data.

The article makes four contributions. First, it synthesises apparently separate oil-palm AI applications as evidence of a common representation problem. Second, it proposes a five-layer conceptual architecture linking multimodal sensing, self-supervised pretraining, plantation representations, task adaptation and operational feedback. Third, it proposes a domain-shift-oriented benchmark in which geographic, temporal, sensor, biological and management transfer are explicit evaluation dimensions. Fourth, it extends foundation-model discussions beyond technical accuracy to data sovereignty, federated learning, explainability, uncertainty and organisational deployment. The ultimate argument is that the scientific test of PalmFM should not be whether it wins one benchmark, but whether a shared representation can reduce labelled-data requirements and maintain performance when the plantation, year, sensor or operational objective changes.

The review is guided by eight research questions. RQ1 asks why existing task-specific AI architectures may be insufficient as a long-term paradigm for oil-palm plantation intelligence, while RQ2 asks which sensory, environmental, agronomic and textual modalities should constitute PalmFM's pretraining information space. RQ3 examines which self-supervised, multimodal and spatiotemporal learning mechanisms are best suited to learning reusable oil-palm representations, and RQ4 asks how tropical domain shift across geography, time, sensors, biology and management should be addressed.

RQ5 asks which downstream tasks provide the strongest test of representation reuse, and RQ6 asks how PalmFM should be benchmarked for transferability, label efficiency, calibration, robustness and operational value. RQ7 examines the governance arrangements required to reconcile cross-company learning with data sovereignty and commercial confidentiality, whereas RQ8 asks how PalmFM could progress from a conceptual architecture to an operational, continually learning plantation-intelligence platform. Together, these questions connect technical representation learning with deployment, evaluation and institutional feasibility.

Literature Review

From task-specific learning to foundation models

The traditional supervised-learning paradigm assumes that the training and deployment data sufficiently represent the same task and statistical environment. This assumption becomes progressively weaker as models move across locations and applications. Remote sensing has consequently become an important environment for transfer learning because the same physical target can be represented differently according to location, season, atmosphere, sensor and spatial scale. Ma et al. (2024) classify several types of domain shift in environmental remote sensing and emphasise that models trained in one domain frequently deteriorate when transferred elsewhere. Domain-adaptation research similarly shows that feature alignment can reduce discrepancies between heterogeneous optical and SAR imagery, while crop-classification research demonstrates the relevance of unsupervised adaptation when labels in the target region are unavailable (Zhang et al., 2022; Kwak and Park, 2022).

Foundation models alter the starting point of this problem. Instead of learning a representation from the limited labels associated with a single downstream problem, the model is first exposed to broad data through pretraining. Wang et al. (2023a) demonstrated the value of remote-sensing-specific pretraining over relying exclusively on natural-image representations. Wang et al. (2023b) further moved the vision-transformer backbone toward remote-sensing foundation-model functionality, while Cha et al. (2024) investigated the performance implications of scaling model size into the billion-parameter range. Multi-task pretraining represents another pathway, attempting to make the learned backbone useful across different remote-sensing objectives rather than optimising it implicitly for a single surrogate task (Wang et al., 2024).

Self-supervision is central because the volume of Earth-observation imagery is far larger than the volume of high-quality annotation. Contrastive learning can encourage representations of related observations to become similar while retaining discriminative information, whereas masked image modelling requires models to reconstruct hidden content and therefore learn spatial-spectral

structure. Remote-sensing research combining contrastive representation learning, masked learning and self-distillation illustrates how a single pretrained backbone can become useful for classification, object detection, segmentation and change detection (Bi et al., 2023; Muhtar et al., 2023). Few-shot studies further suggest that pretrained models can remain useful when downstream labels are scarce, which directly addresses a central constraint in plantation agriculture (Alosaimi et al., 2023; Liu et al., 2024b).

Multimodal and temporal remote-sensing foundation models

A foundation model intended for plantation intelligence cannot treat all imagery as equivalent. Optical sensing records reflected electromagnetic energy and is highly interpretable spectrally but vulnerable to clouds. SAR remains useful under cloud but represents target geometry, moisture and structure through microwave backscatter and therefore requires modality-aware treatment. Hyperspectral information provides narrow spectral discrimination at high dimensionality, while UAV imagery supplies centimetre-level canopy information over smaller areas. Spectral foundation-model research further indicates that spectral observations benefit from domain-specific pretraining rather than being treated as a simple extension of RGB imagery (Hong et al., 2024). Soil, terrain, meteorology and management records exist outside the image domain altogether.

Agricultural studies already show the value of combining modalities. Sentinel SAR-optical fusion improves crop-type mapping because radar and optical signals contribute complementary information, while optical-SAR transformers can learn spatial and temporal crop representations from both streams (Adrian et al., 2021; Li et al., 2022). Synthetic cross-modal approaches have also been investigated as a means of improving classification when observations from one sensor are unavailable (Mirzaei et al., 2023). Multitemporal satellite imagery and spatiotemporal deep-learning architectures demonstrate the wider importance of time for agricultural yield inference (Qiao et al., 2021; Sagan et al., 2021).

The newest foundation-model literature makes these principles substantially more scalable, while recent surveys also identify cross-modal alignment, unified evaluation and security as unresolved challenges for multimodal remote-sensing foundation models (Zhou et al., 2025). SkySense++ uses a factorised design to accommodate information produced by different sensors and improve multimodal Earth-observation transfer (Wu et al., 2025). SUMMIT demonstrates that SAR should not necessarily be treated as merely another image channel; physically relevant properties such as speckle, structural contours and scattering can be incorporated into self-supervised objectives (Du et al., 2025). SeaMo explicitly integrates multimodal and multi-season information, reinforcing the proposition that geospatial representations should include temporal variation rather than relying solely on static snapshots (Li et al., 2026c). Parameter-efficient multimodal adaptation additionally suggests that downstream deployment does not always require full retraining of a very large backbone (Li et al., 2026a; Li et al., 2025).

Most directly relevant is Agri FM, which demonstrates large-scale agricultural foundation-model pretraining using multi-source temporal remote-sensing observations. Its importance for the PalmFM concept lies less in any one downstream result than in its architecture of agricultural transfer: a pretrained representation can support land mapping, field boundaries, agricultural land-cover tasks and crop mapping using common learned features (Li et al., 2026b). Foundation models in smart agriculture are therefore increasingly positioned as an alternative to repeatedly constructing specialised models from scratch (Li et al., 2024b).

Oil-palm AI as a fragmented representation landscape

Oil-palm AI research provides substantial evidence that useful representations already exist, but they are distributed among separate models. Shaharum et al. (2020) demonstrated large-area plantation mapping using cloud-based machine learning. Zheng et al. (2020) addressed cross-regional tree counting through domain adaptation, explicitly showing that regional heterogeneity degrades direct model transfer. Subsequent research used UAV imagery to characterise the growing status of individual palms and explored multisource domain generalisation for VHR satellite-based tree detection (Zheng et al., 2021; Zheng et al., 2022). Object-based deep learning and more recent RetinaNet variants continue to improve individual-tree inventories (Putra and Wijayanto, 2023; Lee et al., 2024).

Health and physiological assessment represent another independent branch. Hyperspectral observations have been used to classify macronutrient status, while proximal multispectral imagery has been used to estimate nutritional conditions in leaves (Amirruddin et al., 2020; Chungcharoen et al., 2022). Basal stem rot research combines hyperspectral UAV sensing, machine learning and deep semantic segmentation to identify disease symptoms (Kurihara et al., 2022; Ong et al., 2023). These models share underlying information concerning canopy colour, spectral response, crown morphology and disease-related structural change, yet these common features are generally relearned separately.

Harvesting and production constitute another silo. Deep learning has been applied to fresh-fruit-bunch ripeness on mobile devices and multi-input grading systems, while yield studies combine remote sensing, weather, soil moisture and management information (Suharjito et al., 2021; Pipitsunthonsan et al., 2023; Watson-Hernández et al., 2022; Khan et al., 2022; Ang et al., 2022; Jamshidi et al., 2024). These problems involve temporal and biological relationships that could potentially benefit from representations learned during plantation-scale pretraining. Recent fruit-counting research and real-time UAV bunch classification further illustrate continued specialisation of individual computer-vision models (Hamdani et al., 2026; Che Hassan et al., 2026).

Structural and sustainability applications broaden the information space again. Oil-palm height can be modelled using ICESat-2 and age information; Sentinel-1 and Sentinel-2 have been combined to distinguish plantations from surrounding forest; and recent deep-learning work has generated regional 10-m time-series plantation maps for Malaysia and Indonesia (Zang et al., 2023; Abramowitz et al., 2023; Kuapanich et al., 2026). Collectively, the literature implies that "oil-palm intelligence" spans individual fruit, palm, block, estate, landscape and regional scales. The absence of a representation system linking those scales is therefore increasingly conspicuous.

Trust, data sharing and operational intelligence

A foundation model trained across plantations would also change the institutional requirements of agricultural AI. Data fusion in agriculture is valuable precisely because individual sources are incomplete or ambiguous, but combining them requires interoperability, metadata consistency and explicit treatment of missing observations (Barbedo, 2022). Agricultural AI also requires interpretability because management recommendations can produce agronomic and economic consequences, and seemingly accurate black-box relationships may not remain reliable under unusual conditions (Ryo, 2022).

Cross-company foundation-model training introduces a further challenge: raw agronomic, yield and operational records may be commercially sensitive. Federated learning provides one possible architecture in which local data remain in participating organisations while model updates are aggregated. Agricultural studies demonstrate the feasibility of cross-silo federated learning and increasingly efficient federated yield models (Durrant et al., 2022; Rizman Žalik and Žalik, 2023; Li et al., 2024a). Agricultural data-space research similarly emphasises interoperability and data sovereignty as design requirements rather than secondary legal concerns (Falcão et al., 2023).

The operational destination of PalmFM should therefore be broader than image interpretation. Environmental applications of foundation models increasingly focus on accessible applications and workflows that translate pretrained representations into user-facing monitoring tools (Strong et al., 2025). Digital-twin research further illustrates how continuously updated sensor data, models and spatial information can be combined to represent changing biological systems (Jiang et al., 2022). These developments suggest that PalmFM could eventually function as the representation backbone of a plantation digital twin rather than as an isolated predictive model.

Method

This article adopts a **qualitative literature review (QLR)** design. It is explicitly **not a systematic literature review** and does not claim exhaustive retrieval, PRISMA-based screening or statistical meta-analysis. The purpose is conceptual integration: to connect research streams that are normally discussed independently and use their intersection to formulate an oil-palm-specific foundation-model architecture.

The review focuses on peer-reviewed journal literature published from 2020 through 2026. Four literature domains were purposively combined: remote-sensing foundation models and self-supervised learning; multimodal and spatiotemporal agricultural AI; oil-palm remote sensing and machine learning; and governance topics including explainability, federated learning, interoperability and data sovereignty. Seminal developments were prioritised where they directly addressed generalisation, multi-task representation, multi-sensor information, label efficiency or real-world plantation applications.

The interpretive procedure followed three analytical stages. First, individual studies were coded according to the problem they addressed: representation learning, modality integration, temporal modelling, transfer, task adaptation, oil-palm operations or governance. Second, related codes were consolidated into higher-order themes. Third, these themes were compared to determine which capabilities would have to coexist in an oil-palm foundation model. The resulting thematic framework contains seven principal dimensions: fragmentation, multimodality, temporality, tropical domain shift, adaptation, evaluation and governance.

The analysis consequently differs from a catalogue-style review. An oil-palm tree detector, for example, is considered not only as an object-detection application but also as evidence regarding which crown-scale representations transfer across regions. A Ganoderma detector is interpreted partly as evidence about spectral and structural health representations. Yield models are interpreted as evidence for environmental and management context. The review therefore asks what reusable knowledge is latent within existing task-specific models and how that knowledge could be integrated during large-scale pretraining.

Because the purpose is interpretive rather than exhaustive, the review does not use frequency counts as evidence of effect size and does not treat absence from the purposively assembled corpus as evidence that a topic is unimportant. Instead, evidence is weighed for conceptual relevance, methodological contribution and transferability to the PalmFM problem. This design supports theory-building across rapidly evolving research streams while making clear that the synthesis is a research architecture rather than a complete census of the literature.

Results: Thematic Findings

Theme 1: Task fragmentation is becoming the principal architectural constraint

The first finding is that contemporary plantation AI possesses breadth without representation integration. Detection models encode crown geometry, nutritional models capture spectral stress, disease models identify abnormal canopy signatures, yield systems model weather-production relationships, and mapping models represent plantation landscape structure. Yet the weights and features learned within these systems generally remain confined to their original problem.

This approach creates at least four costs. First, each downstream task requires new labels. Second, the model repeatedly learns elementary plantation features—crown shape, canopy texture, seasonal reflectance or block structure—that other models have already learned. Third, operational maintenance becomes difficult because an estate may eventually manage dozens of unrelated models. Fourth, improvements from new regional data do not automatically propagate across tasks.

The foundation-model proposition therefore begins with knowledge accumulation. PalmFM should create a persistent representation layer onto which individual applications can be attached. The backbone would absorb large quantities of historical observations even when no task-specific labels are available. Specialised labels would then teach smaller adapters or task heads how to interpret those representations for a particular decision. This changes plantation AI from a portfolio of independently trained models into an accumulating learning system.

Theme 2: The Palm FM information space must be genuinely multimodal

PalmFM should not be synonymous with a satellite-image foundation model. Plantation state is only partially observable through any single sensor. Sentinel-2 captures vegetation-related spectral differences but suffers from tropical cloud contamination. Sentinel-1 provides all-weather observations and information associated with structure and moisture, but radar backscatter is physically different from optical reflectance. VHR commercial satellites reveal crown-scale spatial structure. UAVs provide centimetre-level RGB, multispectral, thermal or hyperspectral information but do not economically cover all estates continuously. Weather stations, reanalysis, soils, elevation and hydrology contextualise remotely sensed changes. Agronomic databases provide planting year, genotype, fertilisation, harvesting and yield information that cannot be reconstructed perfectly from imagery.

A future Palm FM should consequently support four data families: **Earth observation, proximal observation, environmental context,** and **operational knowledge**. The architecture must tolerate missing modalities because every observation will not contain every data type. It must additionally distinguish modality from scale. A Sentinel-2 pixel may describe a 10-m neighbourhood, a UAV image may resolve a frond, and an agronomic record may apply to an entire management block.

Figure 1 summarises the proposed five-layer architecture by showing how heterogeneous plantation observations are transformed into a reusable representation and then into multiple operational applications. It also clarifies where reusable representation learning is separated from task-specific adaptation and operational feedback.

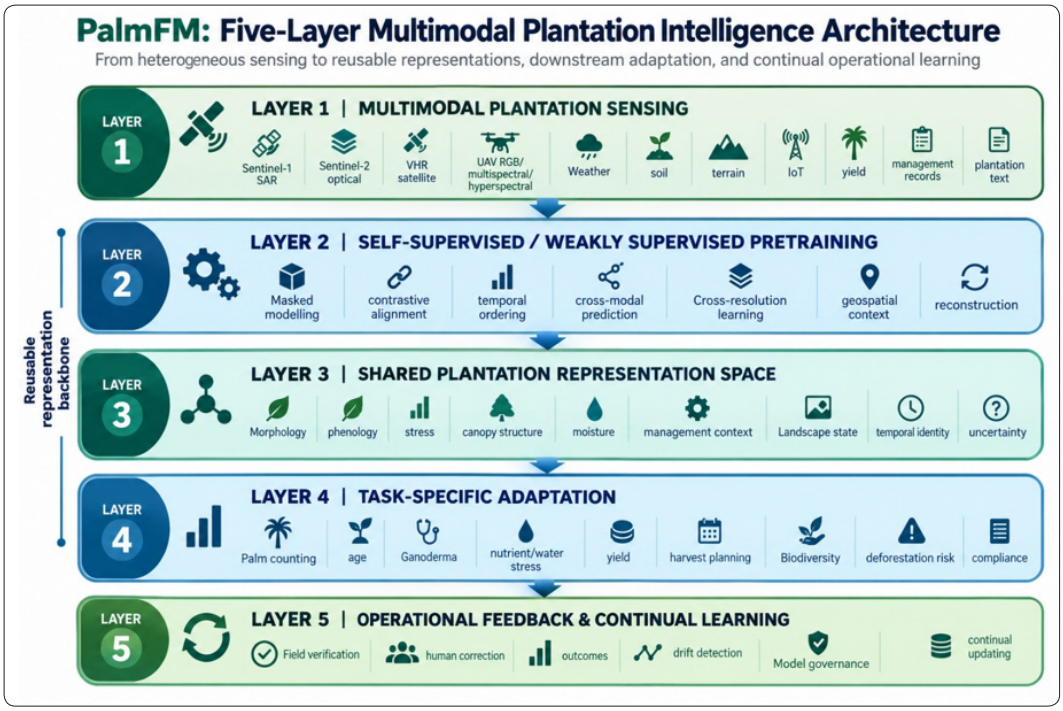


Figure 1: Conceptual five-layer architecture of Palm FM

The essential characteristic of Figure 1 is the separation between representation learning and operational tasks. A conventional model links Layer 1 directly to one item in Layer 4. Palm FM inserts a persistent pretrained representation between them. Consequently, a disease annotation could improve a health-related adapter without requiring the entire plantation-intelligence system to be relearned, while improvements in the shared backbone could benefit several downstream applications.

Theme 3: Pretraining should encode biology, geography and time

Generic masked reconstruction alone is unlikely to capture everything Palm FM needs. Oil-palm pretraining objectives should reflect regularities that are specific to the plantation domain. One objective could reconstruct masked spectral bands or image patches. Another could require the model to identify whether two observations belong to the same palm or block at different times. Cross-modal contrastive learning could align SAR and optical observations from approximately the same place and period. Temporal-order prediction could require the model to infer which observation occurred first. Cross-resolution objectives could align a satellite block representation with aggregated UAV representations.

Phenology presents a particularly important opportunity. A foundation model that has learned normal age- and season-dependent trajectories may identify abnormal behaviour before a supervised disease classifier would do so. The same principle applies to moisture stress, nutrient stress and yield anomaly. Instead of defining every abnormality in advance, Palm FM could learn a manifold of expected plantation trajectories and detect statistically unusual departures.

The temporal dimension also creates a distinction between **state** and **event**. A yellow canopy observed once may be ambiguous. A progressive deviation from the historical trajectory of that palm or block is more informative. PalmFM should therefore preserve temporal identity where feasible: palm, block and estate observations should be linked longitudinally rather than treated as independent images.

Theme 4: Tropical domain shift should become a first-class model object

Domain shift is often reported after a model fails on a new dataset. For Palm FM, it should instead be designed into pretraining and evaluation. Five forms are particularly important. **Geographic shift** occurs when models move between estates, provinces or countries. Soil, rainfall, terrain, cultivar, labour practice and landscape context can all change. **Temporal shift** occurs across seasons and years, including unusual drought, El Niño, flooding or disease cycles. **Sensor shift** occurs when cameras, satellites, spatial resolutions or acquisition geometries change. **Biological shift** arises across planting materials, age classes, canopy structure or health conditions. **Management shift** occurs because fertilisation, drainage, harvesting, pruning and replanting strategies differ among organisations.

Figure 2 represents these shifts as intersecting dimensions rather than independent problems. Figure 2 illustrates why random image-level train-test splitting is insufficient to demonstrate real plantation generalisation.

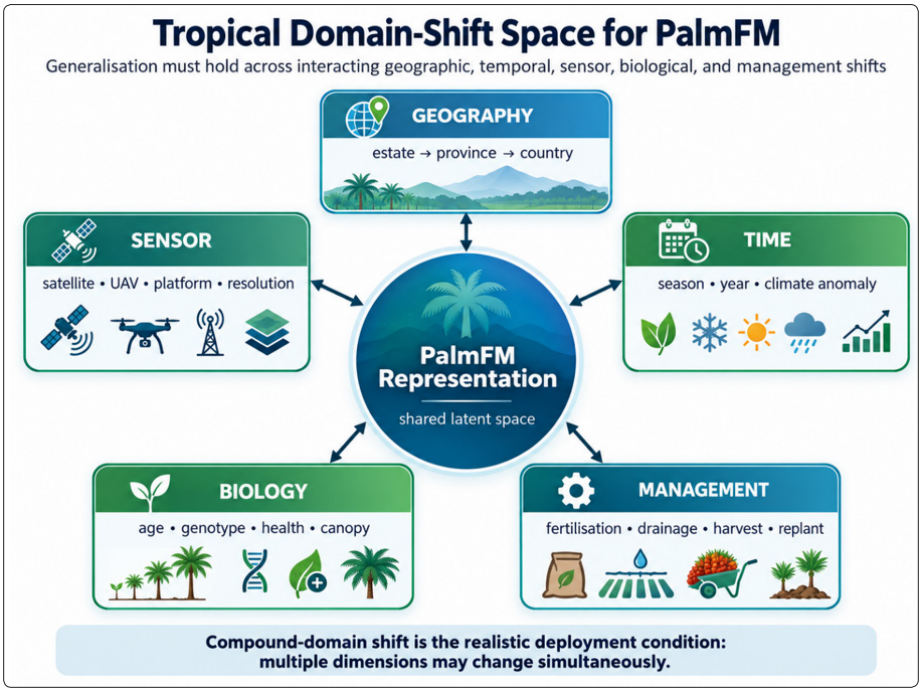


Figure 2: Tropical domain-shift space for Palm FM

The implication of Figure 2 is that operational deployment typically involves several shifts simultaneously. A model may be transferred to another company, observed with a different UAV, during a wetter year and on younger palms. Benchmarking only one perturbation at a time may consequently overestimate robustness. PalmFM evaluation should contain compound-shift scenarios that better approximate real deployments.

Theme 5: Adaptation should test whether the representation is genuinely reusable

A foundation model becomes scientifically interesting only when downstream adaptation is materially more efficient than training specialist models independently. The relevant comparison is therefore not simply final accuracy. PalmFM should be evaluated across a **label-efficiency curve**. At 1%, 5%, 10%, 25% and 100% of available labels, researchers should compare: training a specialist model from scratch; ImageNet-style transfer; general remote-sensing foundation models; agricultural foundation models; and PalmFM.

The downstream benchmark should intentionally mix visual and non-visual tasks. Candidate tasks include individual palm detection and counting, missing-palm identification, age estimation, canopy segmentation, nutrient stress, basal stem rot, water stress, yield forecasting, fresh-fruit-bunch detection, ripeness assessment, harvest planning, replanting prioritisation, plantation boundary extraction, deforestation-risk monitoring, biodiversity proxies and drainage-related hydrological assessment.

This task diversity is critical. A backbone that performs exceptionally on crown segmentation but poorly on temporal yield forecasting may simply be an excellent vision model rather than an oil-palm foundation model. Conversely, successful transfer across structural, physiological, production and environmental tasks would provide stronger evidence that the representation captures plantation-level regularities.

Theme 6: Palm FM requires a new benchmarking philosophy

The most important benchmark design decision is the unit of separation between training and testing. Randomly partitioning nearby image tiles can allow spatial autocorrelation and management similarity to leak into both sets. Consequently, excellent test scores can coexist with poor deployment transfer.

PalmFM should be tested through at least five holdout regimes. **Estate-held-out transfer** trains without any labelled observations from a target estate. **Geography-held-out transfer** removes an entire province or country. **Future-year transfer** trains on previous years and tests on later years, preventing temporal leakage. **Sensor-held-out transfer** evaluates adaptation to a platform absent from pretraining or task-specific training. **Management-held-out transfer** evaluates estates using contrasting operational systems.

Performance metrics should also go beyond mean accuracy. Classification needs class-specific sensitivity and specificity; detection requires precision, recall and counting error; segmentation requires intersection-over-union and boundary quality; regression requires MAE, RMSE and bias. Across all tasks, calibration should be measured because an operational user needs to know whether low-confidence predictions can be identified and escalated for field inspection. This broader evaluation logic is consistent with recent calls for unified, cross-task assessment of multimodal remote-sensing foundation models (Zhou et al., 2025).

Figure 3 integrates pretraining, adaptation and benchmark logic. Figure 3 distinguishes ordinary random-split performance from the more demanding transfer tests required before PalmFM can claim foundation-level plantation generalisation.

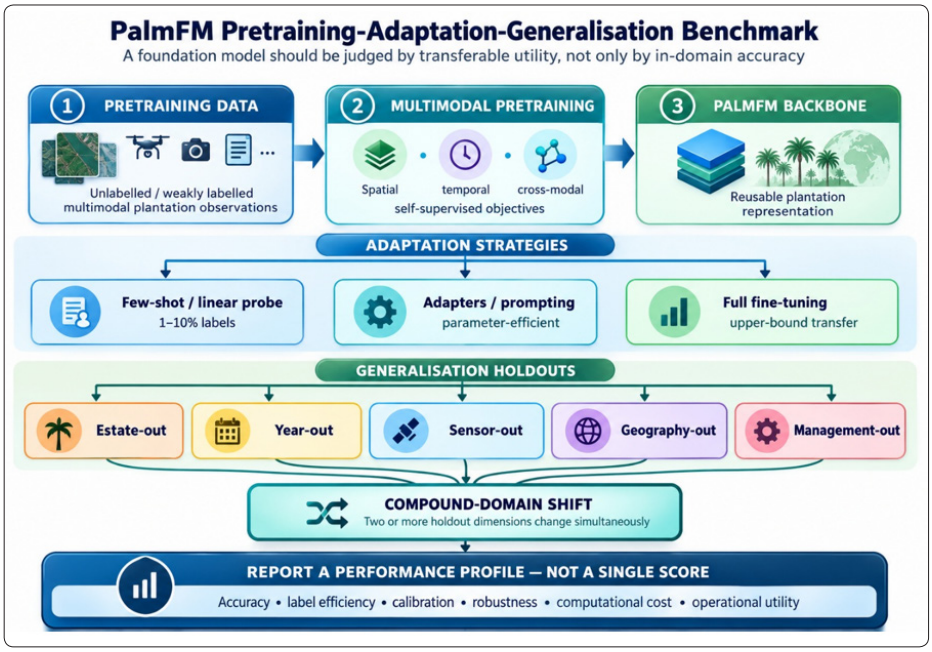


Figure 3: PalmFM pretraining–adaptation–generalisation benchmark

Figure 3 emphasises that benchmark performance should be expressed as a profile rather than one score. A relatively small PalmFM that transfers effectively using few labels and produces calibrated uncertainty could be operationally superior to a much larger model that marginally improves in-domain accuracy but fails when geography or sensor changes.

Theme 7: Data sovereignty determines whether cross-company PalmFM is feasible

The largest potential source of PalmFM performance is also its most difficult institutional problem: diversity across plantations. A model trained on one organisation's history cannot learn the full range of tropical conditions. Yet companies may reasonably refuse to centralise block yields, detailed management histories, maps, cost information or disease records.

Federated pretraining offers one possible compromise. Shared model components could be distributed to participating companies, updated against local data, and aggregated without moving raw records into a common repository. This architecture would not automatically solve privacy or competition concerns—model updates themselves can leak information—but it establishes a technical direction that aligns better with agricultural data sovereignty than unrestricted centralisation.

Figure 4 presents a federated PalmFM arrangement in which companies retain raw data while participating in shared representation learning. The arrangement separates private local training from governed model-update exchange and shared representation learning.

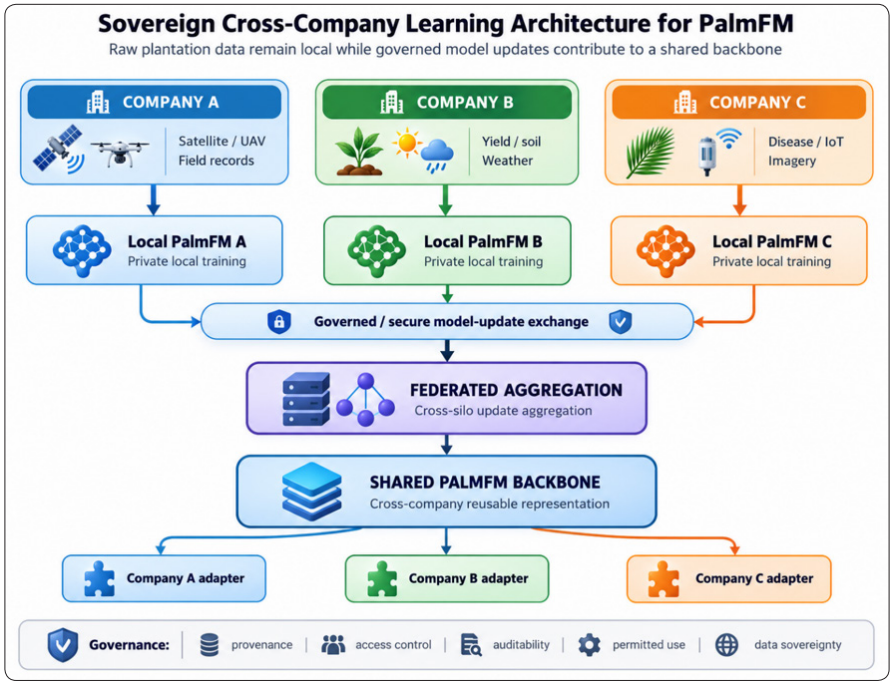


Figure 4: Sovereign cross-company learning architecture for PalmFM

The principal contribution of Figure 4 is organisational rather than algorithmic. A foundation model for an economically strategic commodity cannot assume that all useful information will become open data. Research should therefore evaluate cross-silo learning, access controls, provenance, model licences, auditability and rules governing downstream use from the beginning of the PalmFM programme.

Discussion and Analysis

Why the oil-palm domain is ready for a foundation-model transition

RQ1 asked why task-specific architectures may no longer be sufficient. The literature indicates that the limitation is not poor performance on individual benchmarks. Indeed, specialist models can achieve strong results. The problem is that plantation intelligence now contains enough recurring structure that continually relearning it has become inefficient.

Palm crowns, planting geometry, roads, drainage networks, canopy reflectance, seasonal moisture response, fruit characteristics and age-dependent structure recur across many applications. This suggests that oil-palm AI has moved into a regime where reusable representations may yield increasing returns. Foundation models are attractive precisely because they allow the cost of representation learning to be amortised across tasks. This argument is consistent with agricultural foundation-model research that emphasises reuse, adaptation and label efficiency across multiple downstream tasks (Yin et al., 2025).

However, the transition should not be driven by fashion. If a specialist 20-million-parameter model solves an operational problem reliably with adequate data, replacing it with a billion-parameter architecture provides little intrinsic benefit. PalmFM is justified when shared pretraining demonstrates three outcomes: significantly lower annotation requirements, stronger out-of-domain generalisation and cross-task reuse.

What Palm FM should learn

RQ2 and RQ3 concern modalities and learning mechanisms. The synthesis suggests that PalmFM should learn five categories of latent information : morphology, physiology, environment, phenology and management context.

Morphology includes crown form, spacing, canopy size, palm height and plantation geometry. **Physiology** is represented indirectly through spectral, thermal and structural signals associated with nutrients, disease and water stress. **Environment** incorporates rainfall, temperature, soil, topography, hydrology and surrounding land cover. **Phenology** describes expected temporal development. **Management context** includes planting year, harvesting, fertilisation, drainage, pruning and replanting.

No single sensing modality spans these categories. PalmFM therefore needs modality-specific encoding before alignment in a shared latent space. A practical architecture may use specialised optical, SAR, hyperspectral and tabular encoders combined through attention or learned fusion modules. Missing-modality training should be incorporated deliberately so that inference remains possible where only public Sentinel observations are available. The need for modality-aware encoders is reinforced by spectral and multimodal remote-sensing foundation-model research (Hong et al., 2024; Zhou et al., 2025).

Pretraining should combine multiple losses. Masked reconstruction can develop spatial and spectral understanding; contrastive alignment can bind observations from different sensors; temporal prediction can represent plantation trajectories; cross-resolution alignment can link satellite and UAV representations; and weak agronomic supervision can connect representation learning to biologically meaningful outcomes without transforming pretraining into a narrow supervised task.

Domain shift should define the scientific identity of PalmFM

RQ4 asked how tropical domain shift should be addressed. The central conclusion is that robustness to domain shift should distinguish PalmFM from an ordinary domain-specific vision model.

Indonesia alone contains extensive variation in soils, peat and mineral landscapes, rainfall, plantation age, management systems and surrounding land cover. Expansion to Malaysia, Africa and Latin America introduces further differences in climate, smallholder structure, planting materials and landscape configuration. A PalmFM pretrained exclusively on homogeneous industrial estates could consequently become highly specialised despite its size.

Diversity should therefore be considered a pretraining asset. The corpus should intentionally sample different islands, climatic conditions, age classes, management regimes and sensors. Domain metadata—location, acquisition date, sensor identity, spatial resolution and relevant environmental information—should be retained rather than discarded. These variables can help the model disentangle invariant biological information from domain-specific appearance. Such explicit treatment of domain metadata is aligned with evidence that environmental remote-sensing models can degrade substantially under geographic and sensor distribution shifts (Ma et al., 2024; Zheng et al., 2022).

Evaluation must then reproduce the kinds of deployment shifts that matter. A cross-company benchmark is more informative than a random image split. A future-year drought test is more informative than a conventional validation set. Transfer from Malaysian training sites to Indonesian or African plantations could reveal whether the model has learned true oil-palm structure or only regional visual regularities.

What downstream tasks should be prioritised

RQ5 concerns the most informative downstream tasks. The benchmark should combine tasks with distinct spatial scales and learning requirements. Palm counting provides a clean object-detection test and links naturally to earlier domain-generalisation literature. Ganoderma surveillance and nutrient stress test fine-grained physiological representations. Yield forecasting tests whether temporal imagery and environmental context have been integrated meaningfully. Age and height estimation test structural development. Fruit-bunch detection and ripeness operate at substantially finer spatial resolution and therefore test cross-scale adaptation. Plantation mapping and deforestation monitoring shift the scale outward from palm to landscape.

The strongest demonstration would be multi-task adaptation with intentionally sparse labels. For example, PalmFM could be pretrained without Ganoderma labels, then adapted using disease labels from only a few blocks. If it substantially outperformed generic remote-sensing pretraining and conventional transfer learning under target-estate holdout, this would demonstrate that general plantation representations contain useful health information.

Benchmarking must reward useful transfer rather than benchmark memorisation

RQ6 concerns evaluation. Foundation-model benchmarking frequently risks becoming another leaderboard if all models are repeatedly fine-tuned on established datasets. PalmFM should instead be judged on transfer efficiency.

A recommended reporting framework would include in-domain performance, zero-shot or linear-probe performance where applicable, 1–10% label adaptation, full fine-tuning, cross-domain transfer, calibration, inference cost and energy/computational requirements. The resulting benchmark would discourage the assumption that larger always means better. Recent multimodal foundation-model reviews similarly identify cross-task generalisation and unified evaluation criteria as major open problems (Zhou et al., 2025).

Model size should particularly be treated with caution in plantation environments. Many operational decisions occur on limited-connectivity estates, mobile devices or edge platforms. A very large model might therefore function primarily as a central pretrained backbone from which compressed representations, adapters or student models are derived. Parameter-efficient adaptation and model pruning are relevant because operational implementation can decouple the scale of pretraining from the scale of inference. Agricultural federated-learning research further shows that pruning can reduce model size and communication cost while retaining useful local forecasting performance (Li et al., 2024a).

Governance, sovereignty and cross-company collaboration

RQ7 concerns governance. The foundation-model paradigm makes governance technically consequential because model quality depends on data diversity. If data sharing is blocked entirely, representation diversity suffers. If data centralisation is uncontrolled, organisations may expose commercially sensitive information and lose confidence in the collaboration.

A layered governance system is therefore needed. Public remote-sensing imagery can form a shared baseline. Companies can contribute higher-resolution and agronomic information under controlled arrangements. Federated learning can enable some model improvement without centralising raw data. Secure aggregation and contractual controls can define permitted model uses. A common plantation ontology would help ensure that variables such as yield, disease stage, planting year and management operation are semantically interoperable. This governance direction is consistent with cross-silo federated-learning research in the agri-food sector, which treats distributed model training as a means of improving collaboration without requiring unrestricted centralisation of farm-level data (Durrant et al., 2022; Rizman Žalik and Žalik, 2023).

Data sovereignty should also apply to smallholders. A PalmFM developed exclusively from large industrial plantations may encode management conditions that are inappropriate for smaller farms. External evaluation across smallholder landscapes should therefore be mandatory if the system is subsequently proposed for sector-wide advisory or sustainability monitoring.

A staged Palm FM roadmap

RQ8 asks how the concept could progress toward implementation. A staged programme is preferable to immediately constructing an extremely large multimodal model. PalmFM-Base should begin with abundant public remote-sensing observations. Sentinel-1, Sentinel-2, terrain and weather time series could provide scalable pretraining across Indonesia and Malaysia. Its first benchmarks should include plantation mapping, age-related structure and yield-linked representations.

Palm FM-Multi should introduce VHR, UAV and hyperspectral data. Cross-resolution and missing-modality pretraining would become important at this stage. Disease, nutrition, individual-tree inventory and fruit applications could then be evaluated.

Palm FM-Geo should explicitly optimise domain generalisation. Its corpus should span producing regions beyond Southeast Asia, and benchmarks should require country-held-out transfer and compound domain shifts.

Palm FM-Operations should connect the representation system to estate records, decision workflows and human feedback. Federated learning, audit trails, uncertainty thresholds and continual learning should be introduced at this stage. Continual-learning research in remote sensing also suggests that prompt-based updating can help models incorporate new information while limiting catastrophic forgetting (Zhao et al., 2023).

This progression reduces technical and institutional risk. More importantly, each stage produces a falsifiable research question. If PalmFM-Base does not outperform generic remote-sensing pretraining on oil-palm transfer tasks, the case for subsequent scaling should be reconsidered rather than assumed.

Broader theoretical contribution

Palm FM also provides a broader conceptual lesson for agricultural foundation models. Crop-specific foundation models should not merely be narrower versions of generic geospatial models. Their scientific rationale lies in the possibility that long-lived agricultural systems possess domain regularities worth modelling explicitly.

Oil palm is a perennial crop with long production cycles, stable geolocation and extensive management histories. This creates unusual opportunities for longitudinal representation learning. A model could observe the same palm repeatedly for years, learning its development trajectory and responses to stress and intervention. In contrast with annual crops, the identity of the biological object can persist through a very long time series.

This suggests an eventual transition from **image foundation models** to **plantation-state foundation models**. An image-centric model asks what is visible now. A plantation-state model asks what biological and operational state most plausibly generated the available observations and how that state is changing. Achieving this goal would require additional progress in temporal learning, causal representation, mechanistic crop knowledge and uncertainty quantification, but it offers a more scientifically substantive target than merely scaling a vision transformer.

Limitations and priorities for empirical validation

The PalmFM framework remains a conceptual synthesis rather than an empirically validated foundation model, so its proposed advantages should be treated as testable hypotheses rather than demonstrated performance gains. The review also integrates rapidly evolving literatures whose terminology, architectures and benchmarks are still changing, which can make apparently stable design recommendations provisional. These limitations do not weaken the conceptual case for PalmFM, but they place a premium on transparent benchmarking and external validation.

A second limitation concerns representativeness. Public remote-sensing datasets and published plantation studies do not evenly represent all producing regions, management systems, age

classes or smallholder conditions, and this imbalance could be inherited by any future pretraining corpus. Consequently, PalmFM development should document geographic and organisational coverage explicitly and should evaluate performance separately for industrial estates and heterogeneous smallholder landscapes.

Empirical validation should therefore compare PalmFM against specialist models, general remote-sensing foundation models and agricultural foundation models under identical estate-held-out, geography-held-out, future-year and sensor-held-out protocols. Reporting should include label efficiency, calibration, failure modes, computational requirements and energy cost in addition to conventional predictive accuracy, with compound-domain tests used to approximate realistic deployment conditions (Yin et al., 2025; Zhou et al., 2025). Such evaluation would determine whether oil-palm-specific pretraining provides genuine cross-task and cross-domain value rather than merely increasing model scale.

Conclusion

Artificial intelligence for oil-palm plantations has reached an important architectural turning point. The field now contains increasingly capable models for individual tasks, but continued development through isolated task-specific systems risks creating an expanding collection of models that repeatedly learn similar plantation features, depend on expensive labels and remain vulnerable when transferred across estates, years and sensors. The concept of PalmFM responds to this limitation by treating plantation intelligence as a reusable representation-learning problem. Its proposed architecture combines multimodal sensing, self-supervised and weakly supervised pretraining, a shared plantation representation space, lightweight downstream adaptation and operational feedback. Its defining challenge is tropical domain shift. A credible PalmFM must therefore demonstrate transfer across geography, time, sensors, biological conditions and management systems rather than relying primarily on random image-level validation.

The most promising development strategy is staged rather than monolithic. Initial research should establish whether oil-palm-specific pretraining produces measurable gains over generic remote-sensing and agricultural foundation models under stringent cross-domain tests. Subsequent phases can progressively incorporate UAV, hyperspectral, operational and textual information. Cross-company learning should be designed around interoperability, federated approaches and explicit data sovereignty.The long-term opportunity is consequently larger than building a single large neural network. PalmFM could provide a common representation infrastructure through which future systems for disease surveillance, yield forecasting, precision nutrition, harvesting, biodiversity, compliance and landscape monitoring accumulate knowledge rather than repeatedly starting from zero. If successful, the foundation-model paradigm could shift oil-palm AI from specialised prediction toward continually improving, multimodal and transferable plantation intelligence.

References

1. 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. https://doi.org/10.1016/j.rsase.2023.100968
2. Adrian J, Sidike P, Maimaitijiang M (2021) Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine. ISPRS Journal of Photogrammetry and Remote Sensing 175: 215–235. https://doi.org/10.1016/j.isprsjprs.2021.02.018
3. Alosaimi N, Alhichri H, Bazi Y, Ben Youssef B, Alajlan N (2023) Self-supervised learning for remote sensing scene classification under the few shot scenario. Scientific Reports 13: 433. https://doi.org/10.1038/s41598-022-27313-5
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. https://doi.org/10.1016/j.compag.2020.105768
5. Ang Y, Shafri HZM, Lee YP, et al. (2022) Oil palm yield prediction across blocks from multi-source data using machine learning and deep learning. Earth Science Informatics 15: 2349–2367. https://doi.org/10.1007/s12145-022-00882-9
6. Barbedo JGA (2022) Data fusion in agriculture: resolving ambiguities and closing data gaps. Sensors 22(6): 2285. https://doi.org/10.3390/s22062285
7. Bi M, Wang M, Li Z, Hong D (2023) Vision Transformer with contrastive learning for remote sensing image scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16: 738–749. https://doi.org/10.1109/JSTARS.2022.3230835
8. Bin Mohamad Zaki MA, Ooi J, Ng WPQ, How BS, Lam HL, Foo DCY, Lim CH (2025) Impact of Industry 4.0 technologies on the oil palm industry: a literature review. Smart Agricultural Technology 10: 100685. https://doi.org/10.1016/j.atech.2024.100685
9. Cha K, Seo J, Lee T (2024) A billion-scale foundation model for remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 19: 16363–16378. https://doi.org/10.1109/JSTARS.2024.3401772
10. Che Hassan CAZ, Muhammad Razif MR, Mohd MNH, Abd Halim A, Mohd Mustapha NH, Azman MN (2026) Real-time oil palm fresh fruit bunch drone image classification using YOLO11 architecture. Journal of Oil Palm Research, article in press. https://doi.org/10.21894/jopr.2026.0033
11. Chungcharoen T, Donis-Gonzalez I, Phetpan K, Udompetaikul V, Sirisomboon P, Suwalak R (2022) Machine learning-based prediction of nutritional status in oil palm leaves using proximal multispectral images. Computers and Electronics in Agriculture 198: 107019. https://doi.org/10.1016/j.compag.2022.107019
12. Du Y, Chen Y, Huang L, Yang Y, Ghamisi P, Du Q (2025) SUMMIT: a SAR foundation model with multiple auxiliary tasks enhanced intrinsic characteristics. International Journal of Applied Earth Observation and Geoinformation 141: 104624. https://doi.org/10.1016/j.jag.2025.104624
13. Durrant A, Markovic M, Matthews D, May D, Enright J, Leontidis G (2022) The role of cross-silo federated learning in facilitating data sharing in the agri-food sector. Computers and Electronics in Agriculture 193: 106648. https://doi.org/10.1016/j.compag.2021.106648
14. Falcão R, Matar R, Rauch B, Elberzhager F, Koch M (2023) A reference architecture for enabling interoperability and data sovereignty in the agricultural data space. Information 14(3): 197. https://doi.org/10.3390/info14030197
15. Hamdani H, Septiarini A, Syaifuddin EA, Gading FRN, Winarno E (2026) Automated oil palm fruit counting using deep learning for reliable yield estimation in natural background. Smart Agricultural Technology 14: 102376. https://doi.org/10.1016/j.atech.2026.102376
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. https://doi.

org/10.1109/TPAMI.2024.3362475
17. Huo C, Chen K, Zhang S, et al. (2025) When remote sensing meets foundation model: a survey and beyond. Remote Sensing 17(2): 179. https://doi.org/10.3390/rs17020179
18. Jamshidi EJ, Yusup Y, Hooy CW, Kamaruddin MA, Mat Hassan H, Muhammad SA, Mohd Shafri HZ, 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. https://doi.org/10.1016/j.ecoinf.2024.102595
19. Jiang X, Jiang M, Gou Y, Li Q, Zhou Q (2022) Forestry digital twin with machine learning in Landsat 7 data. Frontiers in Plant Science 13: 916900. https://doi.org/10.3389/fpls.2022.916900
20. Khan N, Kamaruddin MA, et al. (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. https://doi.org/10.3390/plants11131697
21. 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. https://doi.org/10.3390/f17070823
22. Kurihara J, Koo VC, Guey CW, Lee YP, Abidin H (2022) Early detection of basal stem rot disease in oil palm tree using unmanned aerial vehicle-based hyperspectral imaging. Remote Sensing 14(3): 799. https://doi.org/10.3390/rs14030799
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. https://doi.org/10.3390/rs14184639
24. 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. https://doi.org/10.1016/j.compag.2024.109530
25. Li A, Markovic M, Edwards P, Leontidis G (2024a) Model pruning enables localized and efficient federated learning for yield forecasting and data sharing. Expert Systems with Applications 242: 122847. https://doi.org/10.1016/j.eswa.2023.122847
26. Li H, Zheng J, Miao S, Lu R, Cai G, Fu H, Huang J (2026a) Parameter-efficient multimodal adaptation of vision foundation models for remote sensing semantic segmentation. International Journal of Applied Earth Observation and Geoinformation 153: 105520. https://doi.org/10.1016/j.jag.2026.105520
27. Li J, Xu M, Xiang L, Chen D, Zhuang W, Yin X, Li Z (2024b) Foundation models in smart agriculture: basics, opportunities, and challenges. Computers and Electronics in Agriculture 222: 109032. https://doi.org/10.1016/j.compag.2024.109032
28. Li K, Zhao W, Peng R, Ye T (2022) Multi-branch self-learning Vision Transformer (MSViT) for crop type mapping with optical-SAR time-series. Computers and Electronics in Agriculture 203: 107497. https://doi.org/10.1016/j.compag.2022.107497
29. Li W, Liang S, Chen K, Chen Y, Ma H, Xu J, Ma Y, Zhang Y, Guan S, Fang H, Shi Z (2026b) AgriFM: a multi-source temporal remote sensing foundation model for agriculture mapping. Remote Sensing of Environment 334: 115234. https://doi.org/10.1016/j.rse.2026.115234
30. Li X, Li C, Vivone G, Hong D (2026c) SeaMo: a season-aware multimodal foundation model for remote sensing. Information Fusion 125: 103334. https://doi.org/10.1016/j.inffus.2025.103334
31. Li Y, Tan J, Dang B, et al. (2025) Unleashing the potential of remote sensing foundation models via bridging data and computility islands. The Innovation 6(6): 100841. https://doi.org/10.1016/j.xinn.2025.100841
32. Liu F, Chen D, Guan Z, Zhou X, Zhu J, Ye Q, Fu L, Zhou J (2024a) RemoteCLIP: a vision language foundation model for remote sensing. IEEE Transactions on Geoscience and Remote Sensing 62: 1–16. https://doi.org/10.1109/TGRS.2024.3390838
33. Liu F, Zhang T, Dai W, Zhang C, Cai W, Zhou X, Chen D (2024b) Few-shot adaptation of multi-modal foundation models: a survey. Artificial Intelligence Review 57: 268. https://doi.org/10.1007/s10462-024-10915-y
34. Ma Y, Chen S, Ermon S, Lobell DB (2024) Transfer learning in environmental remote sensing. Remote Sensing of Environment 301: 113924. https://doi.org/10.1016/j.rse.2023.113924
35. Mirzaei A, Bagheri H, Khosravi I (2023) Enhancing crop classification accuracy through synthetic SAR-optical data generation using deep learning. ISPRS International Journal of Geo-Information 12(11): 450. https://doi.org/10.3390/ijgi12110450
36. Muhtar D, Zhang X, Xiao P, Li Z, Gu F (2023) CMID: a unified self-supervised learning framework for remote sensing image understanding. IEEE Transactions on Geoscience and Remote Sensing 61: 1–17. https://doi.org/10.1109/TGRS.2023.3268232
37. Ong WK, Tan WC, Tay LC, Lai WK (2023) Early symptom detection of basal stem rot disease in oil palm trees using a deep learning approach on UAV images. Computers and Electronics in Agriculture 213: 108192. https://doi.org/10.1016/j.compag.2023.108192
38. Pipitsunthonsan P, Pan L, Peng S, et al. (2023) Palm bunch grading technique using a multi-input and multi-label convolutional neural network. Computers and Electronics in Agriculture 210: 107864. https://doi.org/10.1016/j.compag.2023.107864
39. Putra YC, Wijayanto AW (2023) Automatic detection and counting of oil palm trees using remote sensing and object-based deep learning. Remote Sensing Applications: Society and Environment 29: 100914. https://doi.org/10.1016/j.rsase.2022.100914
40. Qiao M, He X, Cheng X, Li P, Luo H, Zhang L, Tian Z (2021) Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks. International Journal of Applied Earth Observation and Geoinformation 102: 102436. https://doi.org/10.1016/j.jag.2021.102436
41. Rizman Žalik K, Žalik M (2023) A review of federated learning in agriculture. Sensors 23(23): 9566. https://doi.org/10.3390/s23239566
42. Ryo M (2022) Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. Artificial Intelligence in Agriculture 6: 257–265. https://doi.org/10.1016/j.aiia.2022.11.003
43. Sagan V, Maimaitijiang M, Bhadra S, Maimaitiyiming M, Brown DR, Sidike P, Fritschi FB (2021) Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. ISPRS Journal of Photogrammetry and Remote Sensing 174: 265–281. https://doi.org/10.1016/j.isprsjprs.2021.02.008
44. 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. https://doi.org/10.1016/j.rsase.2020.100287
45. Strong B, Boyda E, Kruse C, Ingold T, Maron M (2025) Digital applications unlock remote sensing AI foundation models for scalable environmental monitoring. Frontiers in Climate 7: 1520242. https://doi.org/10.3389/fclim.2025.1520242
46. 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. https://doi.org/10.1016/j.compag.2021.106359
47. Wang D, Zhang J, Du B, Xia GS, Tao D (2023a) An empirical study of remote sensing pretraining. IEEE Transactions on Geoscience and Remote Sensing 61: 1–20. https://doi.org/10.1109/TGRS.2022.3176603
48. Wang D, Zhang J, Xu M, et al. (2024) MTP: advancing remote sensing foundation model via multi-task pretraining. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17: 11632–11654. https://doi.org/10.1109/JSTARS.2024.3408154
49. Wang D, Zhang Q, Xu Y, Zhang J, Du B, Tao D, Zhang L (2023b) Advancing plain Vision Transformer toward remote sensing foundation model. IEEE Transactions on Geoscience and Remote Sensing 61: 1–15. https://doi.org/10.1109/TGRS.2022.3222818
50. Watson-Hernández F, Gómez-Calderón N, da Silva RP (2022) Oil palm yield estimation based on vegetation and humidity indices generated from satellite images and machine learning techniques. AgriEngineering 4(1): 279–291. https://doi.org/10.3390/agriengineering4010019
51. Wu K, Zhang Y, Ru L, et al. (2025) A semantic-enhanced multi-modal remote sensing foundation model for Earth observation. Nature Machine Intelligence 7: 1235–1249. https://doi.org/10.1038/s42256-025-01078-8
52. 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. https://doi.org/10.3390/rs17142481
53. Yin S, Xi Y, Zhang X, Sun C, Mao Q (2025) Foundation models in agriculture: a comprehensive review. Agriculture 15(8): 847. https://doi.org/10.3390/agriculture15080847
54. 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. https://doi.org/10.1016/j.rse.2023.113693
55. Zhang C, Feng Y, Hu L, Tapete D, Pan L, Liang Z, Cigna F, Yue P (2022) A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images. International Journal of Applied Earth Observation and Geoinformation 109: 102769. https://doi.org/10.1016/j.jag.2022.102769
56. Zhang Z, Hu X, Yang Y, Yang B, Deng K, Dai H, Zhang M (2025) High-quality one-shot interactive segmentation for remote sensing images via hybrid adapter-enhanced foundation models. International Journal of Applied Earth Observation and Geoinformation 139: 104466. https://doi.org/10.1016/j.jag.2025.104466
57. Zhao L, Xu L, Zhao L, et al. (2023) Continual learning for remote sensing image scene classification with prompt learning. IEEE Geoscience and Remote Sensing Letters 20: 1–5. https://doi.org/10.1109/LGRS.2023.3328981
58. Zheng J, Fu H, Li W, Wu W, Yu L, Yuan S, Tao WYW, Pang TK, Kanniah KD (2021) Growing status observation for oil palm trees using unmanned aerial vehicle (UAV) images. ISPRS Journal of Photogrammetry and Remote Sensing 173: 95–121. https://doi.org/10.1016/j.isprsjprs.2021.01.008
59. Zheng J, Fu H, Li W, Wu W, Zhao Y, Dong R, Yu L (2020) Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network. ISPRS Journal of Photogrammetry and Remote Sensing 167: 154–177. https://doi.org/10.1016/j.isprsjprs.2020.07.002
60. Zheng J, Wu W, Yuan S, Fu H, Li W, Yu L (2022) Multisource-domain generalization-based oil palm tree detection using very-high-resolution satellite images. IEEE Geoscience and Remote Sensing Letters 19: 1–5. https://doi.org/10.1109/LGRS.2021.3061726
61. 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. https://doi.org/10.3390/rs17213532.

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.