VOL 4, NO.1 (2024) 16-42
https://jagst.utm.my

JAGST
Journal of Advanced Geospatial Science & Technology

Towards an Annually Updated Oil Palm Age Database (OPAD) with Multimodal Remote Sensing Features: Design and Preliminary Result

Qiang Zhao[1], Le Yu[1,2,3*], Zhenrong Du[4], Kasturi Kanniah[5]
[1]Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Institute for Global Change Studies, Tsinghua University, Beijing 100084, China
[2]Ministry of Education Ecological Field Station for East Asian Migratory Birds, Beijing 100084, China
[3]Tsinghua University (Department of Earth System Science) - Xi'an Institute of Surveying and Mapping Joint Research Center for Next-Generation Smart Mapping, Beijing 100084, China
[4]School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China
[5]Centre for Environmental Sustainability and Water Security (IPASA), Research Institute for Sustainable Environment (RISE) and Tropical Map Research Group, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

*Corresponding author: leyu@tsinghua.edu.cn

Abstract **-** The cultivation of oil palm (*Elaeis guineensis*) is globally significant, particularly in tropical regions, due to the diverse applications of palm oil and palm kernel oil. This necessitates precise monitoring for sustainable management amidst escalating demand. This study aims to develop an annually updated Oil Palm Age Database (OPAD) using multisource remote sensing data to predict the oil palm age, a critical factor affecting productivity and environmental sensitivity. By comprehensively analyzing multisource remote sensing data from 1982 to 2023, this study establishes OPAD and validates its reliability using a random forest classifier. The findings reveal the relationship between remote sensing data features and changes in the oil palm age, offering insights into sustainable management practices. The results indicate an overall accuracy of 86.86 ± 0.92% for grouped age (i.e., seed (1-2 years), young (3-8 years), teen (9-14 years), mature (15-25 years), and post-mature (over 25 years)), with features such as GEDI forest height, ScanSAR HH SAVG, Planet NIR, and Planet Blue being instrumental in age prediction. Integrating optical, synthetic aperture radar (SAR), and canopy height data enables effective age prediction, which is essential for optimizing management strategies, assessing yield potential, and mitigating environmental impacts.

Keywords *– Oil Palm Age Database, Multimodal Remote Sensing, GEDI, SAR, Random Forest*

Article History: Received 14 January 2024, Accepted 25 February 2024, Published 31 March 2024

How to cite: Zhao, Q., Yu, L., Du, Z. and Kanniah, K. (2024). Towards an annually updated oil palm age database (OPAD) with multimodal remote sensing features: design and preliminary result. Journal of Advanced Geospatial Science and Technology. 4(1), 16-42.

1.0 Introduction

The cultivation of oil palm (Elaeis guineensis) has become a pivotal component of the global agricultural landscape, particularly in tropical regions (Cheng *et al.*, 2018; Rizeei *et al.*, 2018; Zhao *et al.*, 2024). Oil palm products, including palm oil and palm kernel oil, are extensively utilized across various industries, ranging from food and cosmetics to biofuels and pharmaceuticals (Li *et al.*, 2018; Mohd Najib *et al.*, 2020; Xu *et al.*, 2020). The substantial economic contributions of the oil palm industry have led to significant expansion of oil palm plantations, especially in Malaysia and Indonesia (Cheng *et al.*, 2016; Xu *et al.*, 2020). As the demand for palm oil continues to rise, there is an increasing emphasis on sustainable oil palm management practices, which require accurate monitoring and assessment of oil palm plantations and their associated vegetation dynamics (Pohl *et al.*, 2016; Shafri *et al.*, 2012).

The age of oil palm trees is crucial in determining their productivity and vulnerability to pests, diseases, and environmental stressors (Hamsa *et al.*, 2019; Tridawati *et al.*, 2018). Understanding the spatial distribution of oil palm age across plantations is vital for optimizing management strategies, assessing yield potential, and mitigating environmental impacts (Chong *et al.*, 2017; Darmawan *et al.*, 2021). Tan *et al.* (2013) identified oil palm age as a significant factor influencing fruit bunch production. However, large-scale predictions of oil palm age using manual sampling methods are inefficient. Remote sensing data has emerged as a critical tool for predicting oil palm age, facilitated by the segmented planting of oil palms in plantations, allowing for studying oil palm age (Chemura *et al.*, 2015; Hamsa *et al.*, 2019).

A substantial body of research has utilized remote sensing data to predict the age of oil palm trees. As documented in prior research, various studies have mapped the oil palm age (Chong *et al.*, 2017; Tan *et al.*, 2013). Commonly employed remote sensing methods in oil palm tree age research include tree crown segmentation using multispectral and high-resolution remote sensing imagery, classification based on vegetation indices and texture features, and the integration of temporal remote sensing data for change detection and growth model establishment (Camacho *et al.*, 2019; Mohd Najib *et al.*, 2020).

The intricate relationship between oil palm age and spectral bands in optical remote sensing data provides insights into the growth status, health conditions, and ecological influences on oil palm vegetation (Mohd Najib *et al.*, 2020). Researchers have employed optical band ratios or biophysical parameters of oil palm, such as leaf area, canopy cover, and stem height, fitted against oil palm age, selecting indicators with solid correlations to estimate oil palm age (Chemura *et al.*, 2015; McMorrow, 2001). For instance, the infrared bands primarily reflect vegetation’s water content and structure. With increasing oil palm age, vegetation water content and structure variations may influence the infrared bands' reflectance (Tridawati *et al.*, 2018). Optical remote sensing data can also characterize vegetation growth and health conditions through the computation of spectral indices, such as the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), which can reflect vegetation greenness and vigour, potentially correlate with oil palm age. Sitorus. (2010) found a significant correlation between Landsat TM bands and oil palm age, especially in the B5 (shortwave infrared 1) band, as well as with the Infra-Red index (IRI) and Middle Infra Red Index (MIRI).

Estimating the age of oil palm trees using Synthetic Aperture Radar (SAR) data is considered superior to optical remote sensing data. SAR, an active remote sensing technology, can acquire surface information under various ground and weather conditions. During the growth process of oil palm trees, their structure and physical characteristics change with age, which is reflected in the bands and texture features of SAR data (Darmawan *et al.*, 2021; Mohd Najib *et al.*, 2019). Different bands of SAR data reflect different surface characteristics. For example, VV-polarized images are generally more sensitive to the vertical components of vegetation structure, while HH-polarized images are better at capturing horizontal surface features. As the oil palm age increases, its structure becomes more complex, leading to changes in the reflection characteristics of SAR data bands.

Additionally, the texture features of SAR data are closely related to oil palm age (Avtar *et al.*, 2013; Hamsa *et al.*, 2019; Tan *et al.*, 2013). Changes in the morphology, density, and distribution of the canopy during the growth process of oil palm age affect the texture features of SAR data. Avtar *et al.* (2013) utilized multi-frequency, multi-polarization SAR data to monitor oil palm growth stages in Sarawak, Malaysia, finding that PALSAR data with HV polarization exhibited the highest sensitivity to oil palm age. By analyzing the texture features of SAR data, the distribution and growth status of oil palm age can be inferred.

This study will comprehensively analyze oil palm age distribution and related vegetation characteristics using multisource remote sensing data. By integrating literature data and remote sensing information on the oil palm age from 1982 to 2023, an annually updated oil palm age database (OPAD) is generated. Utilizing Google Earth Engine enables the extraction of multisource remote sensing data, including optical, SAR, and canopy height information, facilitating detailed analysis of remote sensing data features for different age groups of oil palm. This study provides a basis for sustainable oil palm management practices and land use planning by elucidating the multisource remote sensing information characteristics of oil palm age variations. The reliability of OPAD is evaluated using a random forest classifier to ensure its suitability for further analysis and application.

2.0 Materials and Methods

2.1 Annually updated oil palm age database (OPAD)

We compiled oil palm age data from multiple studies and calibrated it using high-resolution remote sensing data to ensure accuracy. The goal was to assemble a comprehensive dataset spanning various years and geographical regions, encompassing a wide range of oil palm age categories. As detailed in Table 1, we obtained 1041 oil palm samples from six plantations across four studies. We extended these samples into long-term time series by correlating them with high-resolution satellite imagery from Google Earth, identifying the initial year of oil palm planting at each sample point and any subsequent replanting events. We constructed the OPAD, which is annually updated from 1982 to 2023.

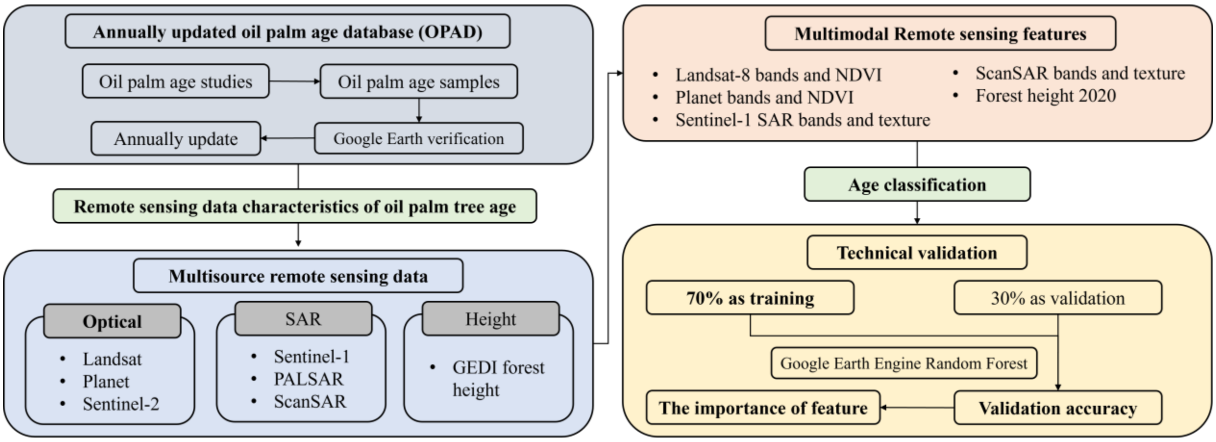

Figure 1. Flowchart of establishing annually updated oil palm age database (OPAD)

Table 1. References on the age of oil palm trees

Author_Location	Count	Reference
Hamsa_Tereh Selatan estate-MAS	167	(Hamsa *et al.*, 2019)
Carolita_Asahan-IDN	206	(Carolita *et al.*, 2019)
Tan_Genting Kulai Besar Estate-MAS	42	(Tan *et al.*, 2014)
Tan_MPOB Kluang-MAS	195	(Tan *et al.*, 2014)
Tan_Sime Darby estate-MAS	322	(Tan *et al.*, 2014)
Toh_Padang Lekir-MAS	109	(Toh *et al.*, 2019)
Total	1041	

As illustrated in Figure 2, the samples we collected are spread across the Malay Peninsula and Sumatra Island. Five of the six sampling sites are located on the Malay Peninsula, contributing 835 samples. Additionally, one site on Sumatra Island comprises 206 samples. In the studies conducted by Hamsa *et al.* (2019) and Carolita *et al.* (2019), age data were available for entire blocks within the plantations. Consequently, we manually collected samples within these blocks to accurately position them on the crowns of oil palm trees. However, it is essential to acknowledge that potential discrepancies in Google Earth imagery may affect the exactitude of these positions. In contrast, the studies by Tan *et al.* (2014) and Toh *et al.* (2019) provided precise field sampling data, with each sample point representing an individual oil palm tree collected directly from the field. These field-based samples offer a higher degree of accuracy in determining the age of oil palms.

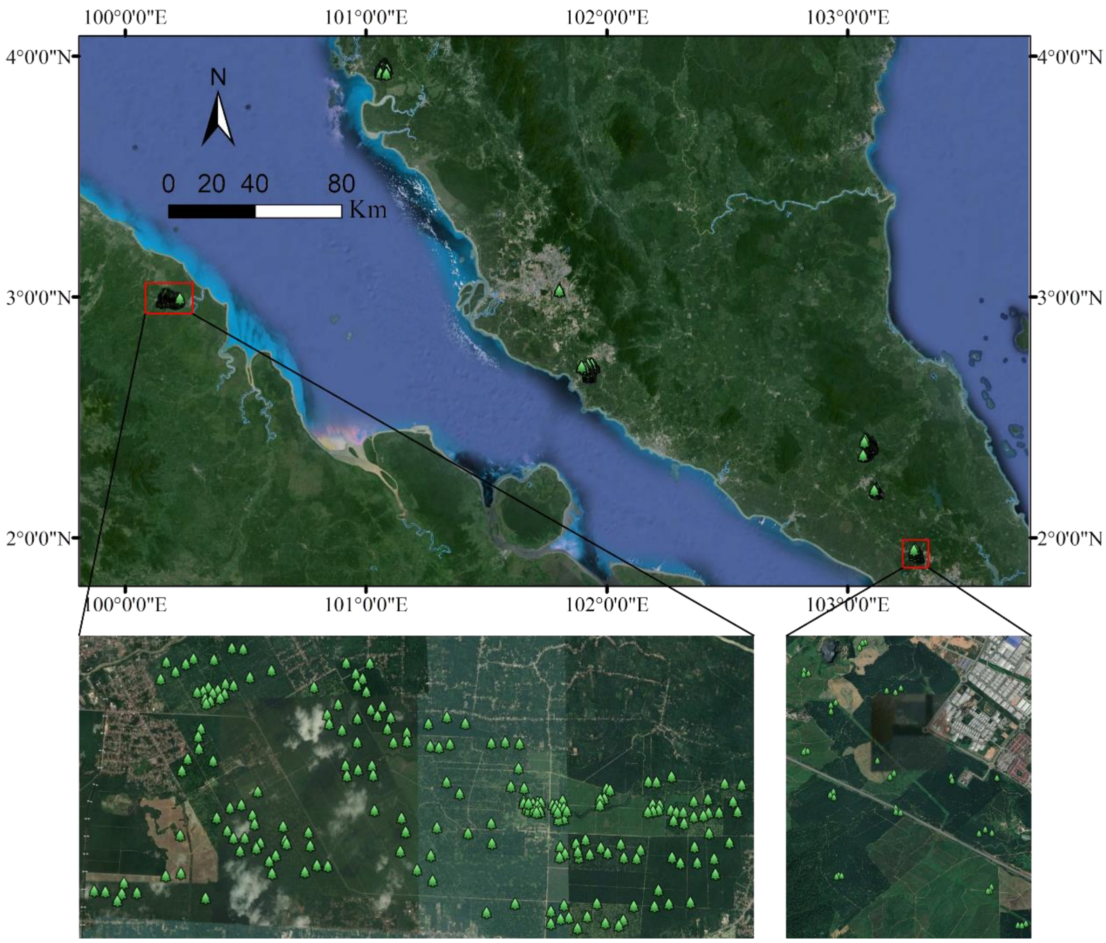

Figure 2. Samples of oil palm tree ages

2.2 Multisource remote sensing data

To investigate the efficacy of various remote sensing data types in accurately determining oil palm age and to elucidate the correlation between these data features and changes in oil palm age, we have chosen a comprehensive set of seven remote sensing data types. This selection encompasses optical, SAR, and canopy height data, as detailed in Table 2.

Table 2. Multisource remote sensing data used for studying the relationship with oil palm age

Remote sensing data		**Source**
Optical	Landsat (Landsat-5 TM, Landsat-7 ETM+, Landsat-8 OLI), (since 1972)	https://developers.google.com/earth-engine/datasets/catalog/landsat
	Planet-NICFI mosaics (since 2015)	https://developers.google.com/earth-engine/datasets/catalog/projects_planet-nicfi_assets_basemaps_asia
	Harmonized Sentinel-2 MSI (since 2017)	https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED
SAR	Sentinel-1 SAR GRD (since 2014)	https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD
	Global PALSAR-2/PALSAR Yearly Mosaic (2007-2010, 2015-2020)	https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_PALSAR_YEARLY_SAR
	PALSAR-2 ScanSAR Level 2.2 (since 2014)	https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_PALSAR-2_Level2_2_ScanSAR
GEDI forest height	Forest Extent and Height Change (2000-2020)	https://glad.umd.edu/dataset/GLCLUC2020

The optical data utilized include Landsat (Landsat-5 TM [Thematic Mapper], Landsat-7 ETM+ [Enhanced Thematic Mapper Plus], Landsat-8 OLI [Operational Land Imager]), Planet-NICFI [Norwegian International Climate and Forest Initiative] mosaics, and Harmonized Sentinel-2 MSI (Multispectral Instrument). SAR data employed consist of Sentinel-1 SAR GRD (Ground Range Detected), Global PALSAR-2/PALSAR (Phased Array type L-band Synthetic Aperture Radar) Yearly Mosaic, and PALSAR-2 ScanSAR Level data. PALSAR-2 ScanSAR data pertains to PALSAR-2 Wide Observation Mode normalized backscatter data, with an observation width of 350 km and a repeat frequency of approximately 42 days (Rosenqvist *et al.*, 2014). Oil palm tree height is sourced from GEDI (Global Ecosystem Dynamics Investigation) forest height data.

To accurately assess oil palm age and explore the relationship between remote sensing data features and age variations, we selected seven types of remote sensing data, including optical, SAR, and canopy height data. We enhanced the precision of OPAD by extending the multisource remote sensing data to represent its characteristics better. We considered different spectral bands in optical remote sensing data and calculated the corresponding Normalized Difference Vegetation Index (NDVI) using the red and near-infrared bands. For SAR data, we examined various polarization bands and computed grey-level co-occurrence matrices for different bands to investigate the texture features of oil palms at different ages. We identified seven texture features used in the study: Difference Entropy (DENT), Dissimilarity (DISS), Inverse Difference Moment (IDM), Information Measure of Correlation 1 (IMCORR1), Information Measure of Correlation 2 (IMCORR2), Sum Average (SAVG), Sum Entropy (SENT).

2.3 Remote sensing data features for tree age classification

To ascertain the reliability of the OPAD, we employed random forests as our validation methodology. This method served dual purposes: it validated the precision of our dataset and quantitatively evaluated the correlation strength between the various remote sensing features and the age of oil palms. The dataset was partitioned, allocating 70% of the oil palm samples for training and reserving 30% for validation following the procedures detailed in Section 2.3. We conducted a total of 12 data validation sets, with each set undergoing ten repetitions to ensure the precision of our results. The average error and standard error for each validation set were meticulously calculated.

Table 3 presents our analysis incorporating a robust suite of 46 remote sensing features. This collection included various spectral bands from optical data, complete with their respective

NDVI calculations. We also included data from multiple polarization bands of SAR, enriched with texture features computed from these bands. We integrated oil palm tree height data from the Global Ecosystem Dynamics Investigation (GEDI) Height dataset to enhance our dataset. Following the validation of the oil palm age data's accuracy, we proceeded to rank the significance of each remote sensing feature to oil palm age prediction.

Table 3. The multimodal remote sensing features

Data Source	Bands	Index or Texture
Landsat-8	B1: Ultra blue	
	B2: Blue	
	B3: Green	
	B4: Red	NDVI
	B5: Near-infrared	
	B6: Shortwave infrared 1	
	B7: Shortwave infrared 2	
Planet-NICFI	B: Blue	
	G: Green	
	R: Red	NDVI
	N: Near-infrared	
Sentinel-1 SAR GRD	VV: Single co-polarization, vertical transmit/vertical receive	DENT, DISS, IDM, IMCORR1, IMCORR2, SAVG, SENT
	VH: Dual-band cross-polarization, vertical transmit/horizontal receive	DENT, DISS, IDM, IMCORR1, IMCORR2, SAVG, SENT
PALSAR-2 ScanSAR	HH: Single co-polarization, horizontal transmit/horizontal receive	DENT, DISS, IDM, IMCORR1, IMCORR2, SAVG, SENT
	HV: Dual-band cross-polarization, horizontal transmit/vertical receive	DENT, DISS, IDM, IMCORR1, IMCORR2, SAVG, SENT
GEDI forest height	Forest height 2020	

Note: DENT: Difference Entropy, DISS: Dissimilarity, IDM: Inverse Difference Moment, IMCORR1: Information Measure of Correlation 1, IMCORR2: Information Measure of Correlation 2, SAVG: Sum Average, SENT: Sum Entropy.

Given the varying temporal coverage of the feature datasets, the validation of oil palm age data and the extraction of feature importance were stratified into three categories. Category 1 (C1) encompassed Landsat-8, PALSAR-2 ScanSAR (ScanSAR), and Sentinel-1 SAR GRD (Sentinel-1) data, covering the period from 2014 to 2023. Category 2 (C2) expanded upon this by incorporating Planet-NICFI (Planet) data features, extending the timeframe from 2015 to

2023. Finally, Category 3 (C3) integrated Forest Height 2020 data into Category 2, focusing the temporal scope to 2020.

In light of the well-defined growth stages observed in oil palm development (Fitrianto *et al.*, 2018), we conducted a detailed validation of OPAD's accuracy across the various growth stages of oil palms. These stages were delineated into five distinct categories (grouped age): Seed (1-2 years), Young (3-8 years), Teen (9-14 years), Mature (15-25 years), and Post-mature (over 25 years) (Chong *et al.*, 2017; Fitrianto *et al.*, 2018). Furthermore, considering the heterogeneity of the collected samples, we bifurcated the OPAD validation process into two approaches: Interpreted Sample validation, which included data from Hamsa_Tereh Selatan estate-MAS and Carolita_Asahan-IDN, and Field Sample validation, which involved data from Tan_Genting Kulai Besar Estate-MAS, Tan_MPOB Kluang-MAS, Tan_Sime Darby estate-MAS, and Toh_Padang Lekir-MAS. Following a thorough analysis of the oil palm age database and feature data, we performed validation on 12 distinct age data sets and ranked the importance of six feature data sets.

3.0 Results

3.1 Optical remote sensing data characteristics of oil palm tree age

The Planet satellite fleet offers high-fidelity Earth observation data that is pivotal for environmental monitoring, agricultural oversight, urban development, and disaster management applications. Figure 3 illustrates a robust correlation between the spectral bands captured by Planet satellites and the age of oil palms. The quartet of Planet satellite bands demonstrates heightened values during the seed stage of oil palm cultivation, which then taper off during the early stage, eventually reaching a steady state. The Near-Infrared (NIR) band, in particular, exhibits more pronounced oscillations, signifying the diverse developmental stages of oil palms. The NIR values are relatively elevated in the Seed stage, subside in the Young stage, stabilize through the Teen and Mature stages, and escalate once more in the Post-mature stage. This underscores the Planet satellite data's critical role as a foundational resource in our investigation, facilitating a nuanced understanding of the interplay between the oil palm age and a multifaceted array of remote sensing features.

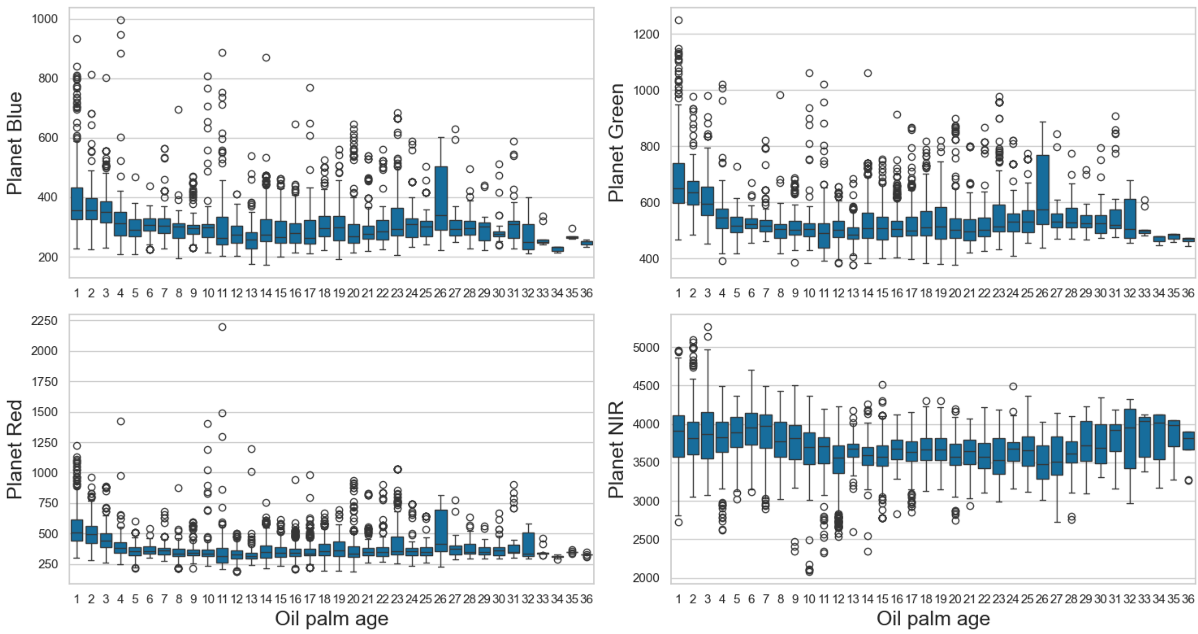

Figure 3. Relationship between bands of Planet imagery and oil palm age (Scale: 0.0001)

Figure 4 delineates the NDVI characteristics of optical remote sensing data from various sources, including Landsat-5, Landsat-7, Landsat-8, Planet, and Sentinel-2, under different age groupings of oil palms. The NDVI is a widely recognized index in remote sensing employed to gauge the extent and vigour of green vegetation on Earth's surface. As the oil palm age advances, there is a concomitant increase in vegetation cover and growth conditions, reflected in rising NDVI values (Tridawati *et al.*, 2018). Observations from Figure 4 reveal that except for Landsat-5, the NDVI values for all other optical remote sensing datasets exhibit an initial increase during the early growth stages of oil palms, followed by stabilization. This pattern is attributed to the rapid growth rate of oil palms during the Seed period, which correlates with an expansion of vegetation cover and a swift rise in NDVI values.

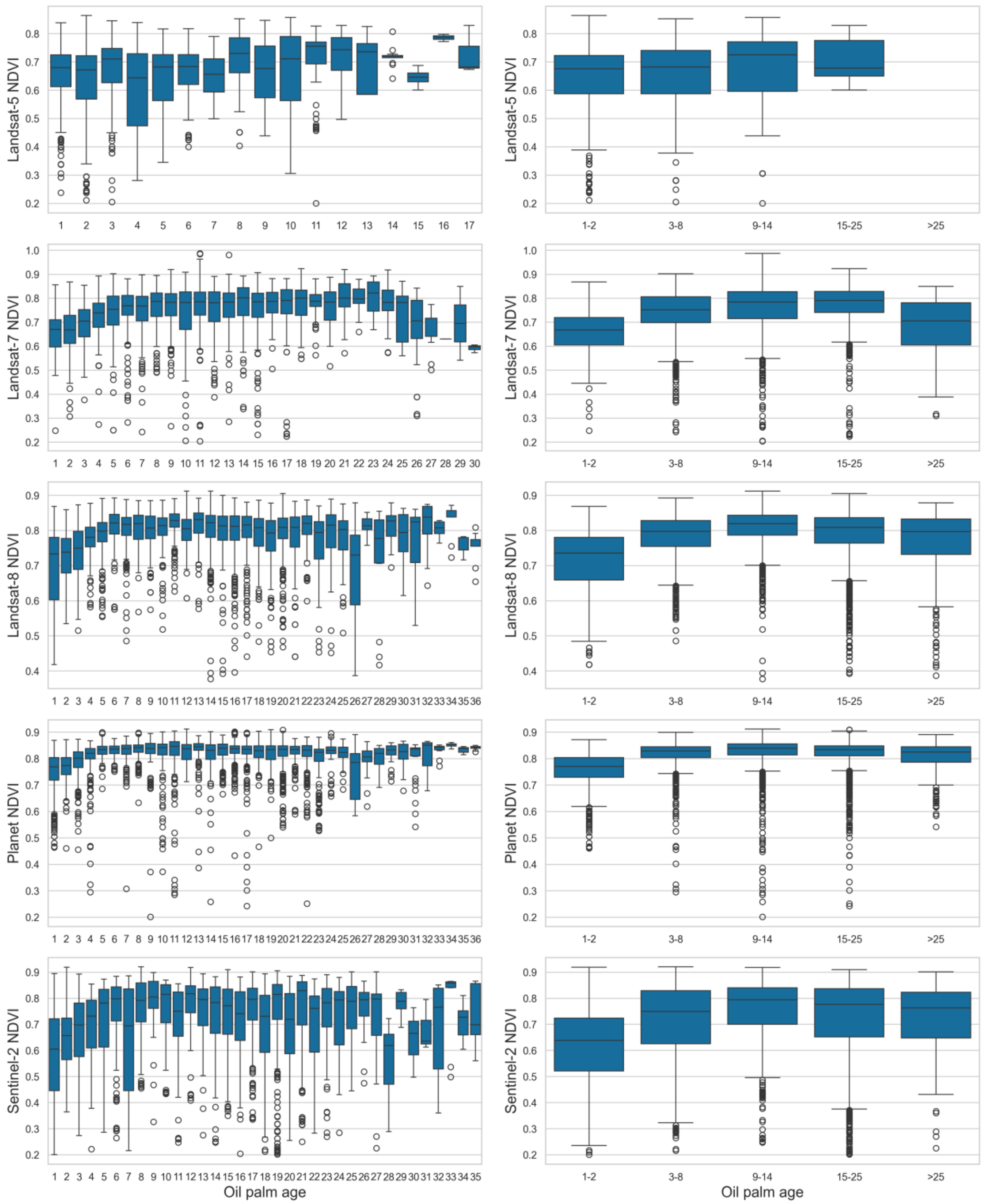

Figure 4. The NDVI features of optical remote sensing data for oil palm tree age

3.2 SAR remote sensing data characteristics of oil palm tree age

3.2.1 Sentinel-1 SAR GRD

By meticulously examining the bands and textural attributes of Synthetic Aperture Radar (SAR) data, we can deduce the distribution and developmental status of oil palm age groups. To this end, we have scrutinized the textural features of Sentinel-1, PALSAR-2/PALSAR (PALSAR), and ScanSAR, alongside their respective bands, to capture the SAR remote sensing data's representation of oil palm across various age stages. Following initial observations and a rigorous screening process, we have identified seven textural features for detailed discussion. This will serve as a foundation for subsequent data validation and analysis of feature importance.

Figure 5 highlights that the Sentinel-1 VV band and its associated textures undergo discernible changes with the age of oil palms. Notably, the VV band parallels the trends observed in optical remote sensing data, with the backscattering coefficient escalating during the early growth stages (Seed and Young periods) and reaching a plateau as the palms mature. Among the textural features, SAVG, DENT, IDM, and DISS are effective indicators of the oil palm's growth trajectory, particularly before age 11, exhibiting marked correlations. SAVG, in particular, maintains an upward trend between the ages of 11 and 16, which indicates its strong relationship with oil palm age. Conversely, IMCORR1 and IMCORR2 do not display as pronounced a relationship, and their correlation with oil palm age is less evident.

Figure 6 presents the variation in the Sentinel-1 VH band and its textural characteristics to the age of oil palms. In contrast to the VV band, the VH band does not display substantial changes in the backscattering coefficient or its textural features across different oil palm age groups. However, a pattern of change is still perceptible. Despite the modest fluctuations in the VH band's backscattering coefficient, subtle differences in textural values, particularly SAVG, DENT, IDM, and DISS, are identifiable, especially for palms younger than 11 years. Similarly, the IMCORR1 and IMCORR2 textural features maintain a relatively stable profile, lacking distinct patterns that could be linked to the age of oil palms. Consequently, it is reasonable to conclude that these two textural features have a limited correlation with the age of oil palms.

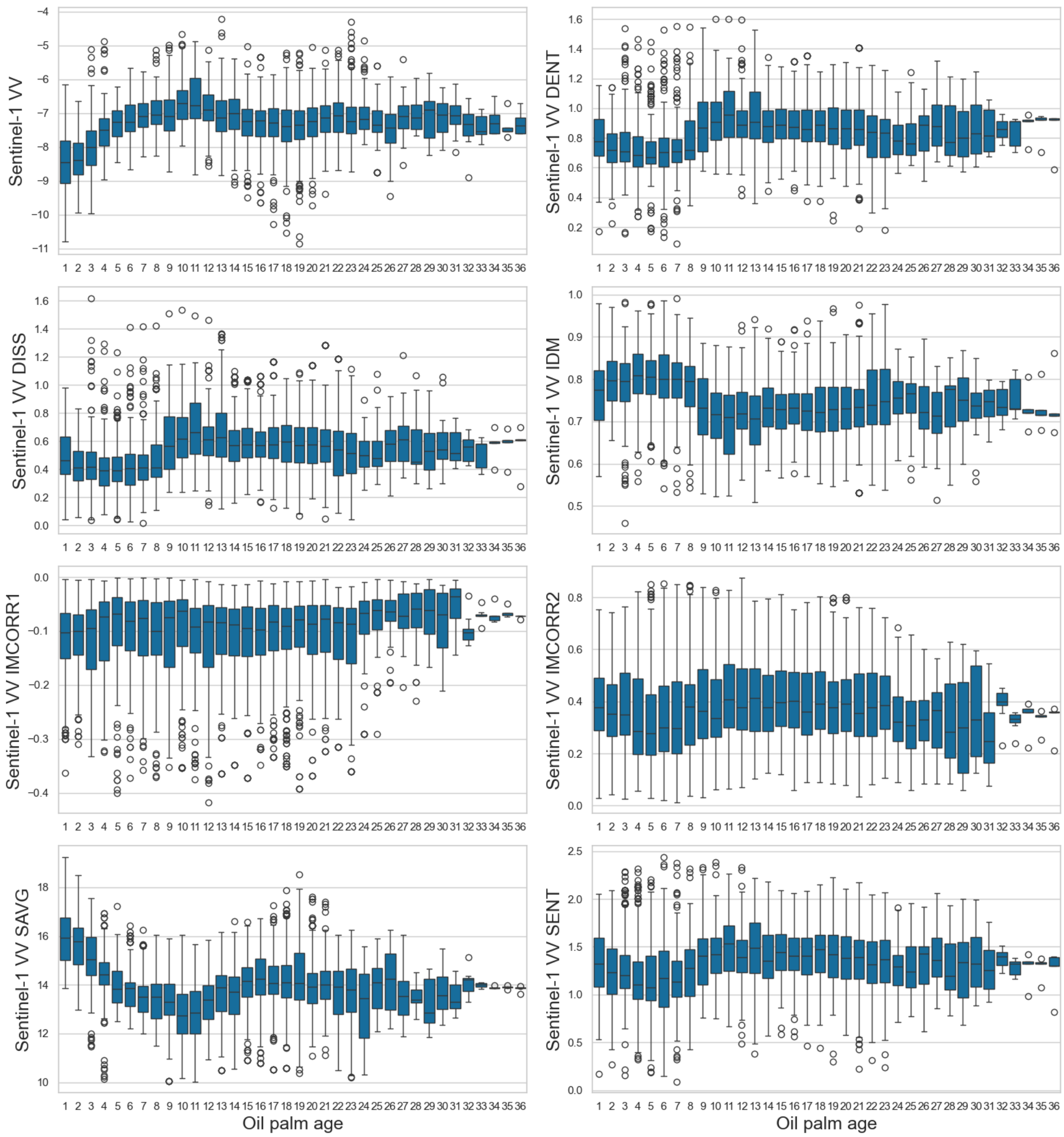

Figure 5. Relationship between Sentinel-1 VV and texture features with oil palm age (absolute value of VV backscattering coefficient considered during texture calculation)

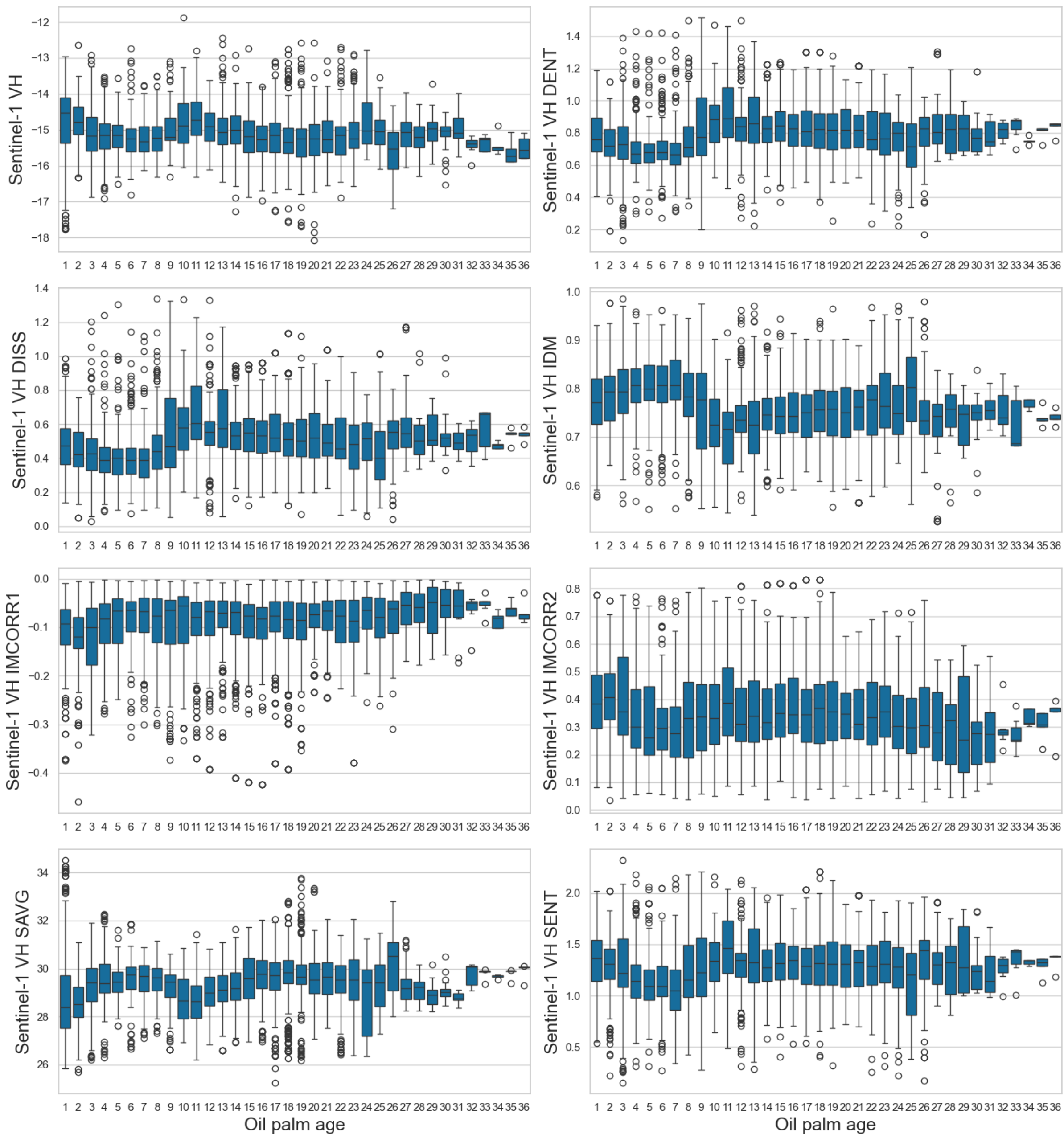

Figure 6. Relationship between Sentinel-1 VH and texture features with oil palm age (absolute value of VH backscattering coefficient considered during texture calculation)

3.2.2 Global PALSAR-2/PALSAR Yearly Mosaic

The fundamental difference between PALSAR and Sentinel-1 is rooted in their respective operational wavelengths. Sentinel-1 functions in the C-band, characterized by medium wavelengths, whereas PALSAR operates in the L-band, which encompasses longer wavelengths. The C-band SAR is known for its enhanced spatial resolution and heightened sensitivity to minute alterations in the Earth’s surface microstructure and texture. This sensitivity, coupled with its reduced interference from vegetation, allows it to detect nuanced differences in oil palm crowns, thereby enabling more accurate differentiation between oil palms of varying ages. On the other hand, the L-band SAR, utilized by PALSAR, boasts superior penetration capabilities, which can access deeper subsurface information. Considering the substantial root systems of oil palm trees, PALSAR remains a valuable asset in studying the oil palm age, offering insights that are not readily accessible through C-band SAR.

As depicted in Figure 7, the backscattering coefficient of the PALSAR HH band exhibits a consistent increase with oil palm age, particularly accelerating during the Seed and Young periods before gradually stabilizing. Similarly, SAVG demonstrates a trend parallel to that of the HH band. Furthermore, DENT, IDM, and DISS all peak during the early growth stages of oil palm, typically around six years of age. Conversely, IMCORR1 and IMCORR2 show relatively minor fluctuations with oil palm age. Figure 7 illustrates that the backscattering coefficient of the PALSAR HH band shows a steady rise in correlation with the age of oil palms, with the rate of increase being most pronounced during the Seed and Young stages before it levels off. This pattern is mirrored by the SAVG feature, which also follows a similar trajectory. Additionally, the textural features DENT, IDM (Inverse Difference Moment), and DISS peak during the early growth stages of oil palms, typically around the age of six years. In contrast, IMCORR1 and IMCORR2 exhibit only marginal variations in the age of oil palms.

Figure 8 presents the backscattering coefficient and textural features associated with the PALSAR HV band. Notably, the HV band of PALSAR reveals patterns analogous to those observed in the HH band. Specifically, IMCORR1 and IMCORR2 exhibit more distinct variations, echoing the DENT, IDM, and DISS trends. This suggests that the textural features of IMCORR1 and IMCORR2 within the PALSAR HV band are more pivotal for the study of oil palm age, surpassing the significance of similar features in the PALSAR HH band and the Sentinel-1 VV and VH bands.

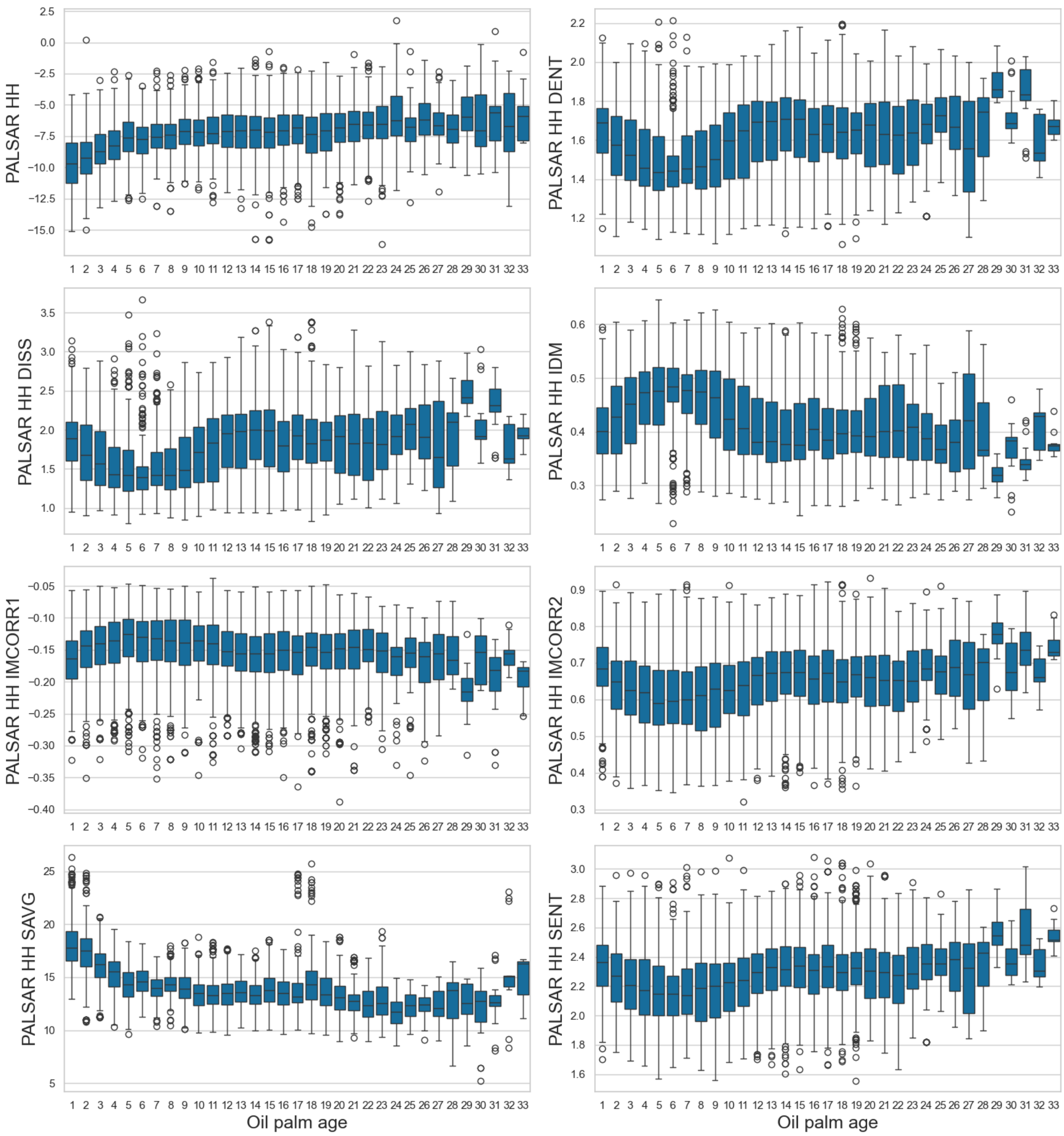

Figure 7. Relationship between PALSAR HH and texture features with oil palm age (absolute value of HH backscattering coefficient considered during texture calculation)

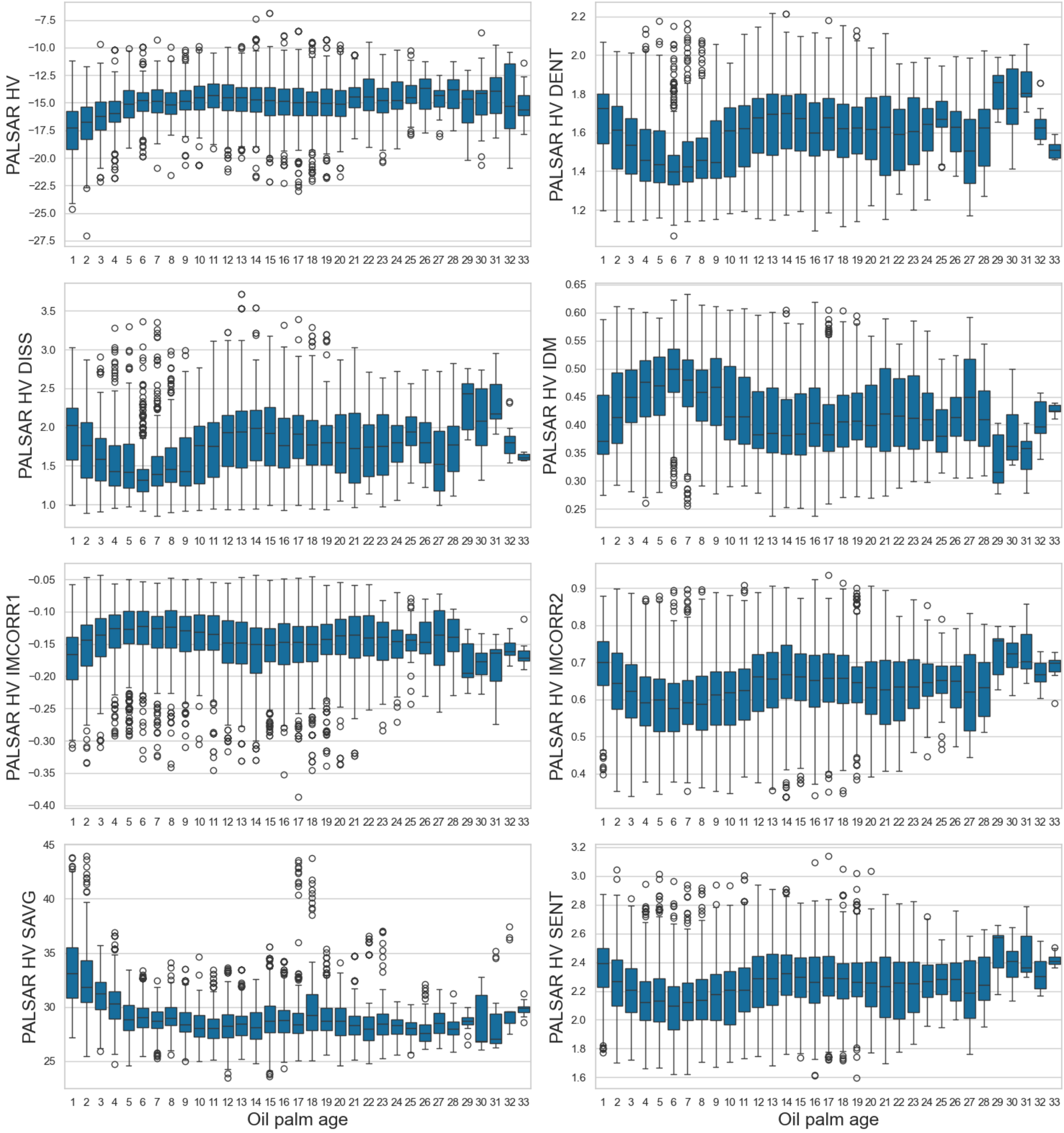

Figure 8. Relationship between PALSAR HV and texture features with oil palm age (absolute value of HV backscattering coefficient considered during texture calculation)

3.2.3 PALSAR-2 ScanSAR

The PALSAR-2 ScanSAR mode facilitates the capture of high-fidelity radar imagery across extensive areas, enhancing the comprehensiveness of surface information obtained. Figures 9 and 10 showcase the backscattering coefficients and textural features of the HH and HV bands operating in ScanSAR. These figures reveal that the HH and HV bands in ScanSAR display trends analogous to those observed with PALSAR, to a certain degree mirroring the developmental characteristics of oil palms across various age stages. Nonetheless, for additional textural features like DENT, IDM, and DISS, the ScanSAR mode does not exhibit markedly distinct features, suggesting a more nuanced interpretation may be required for these specific textural attributes.

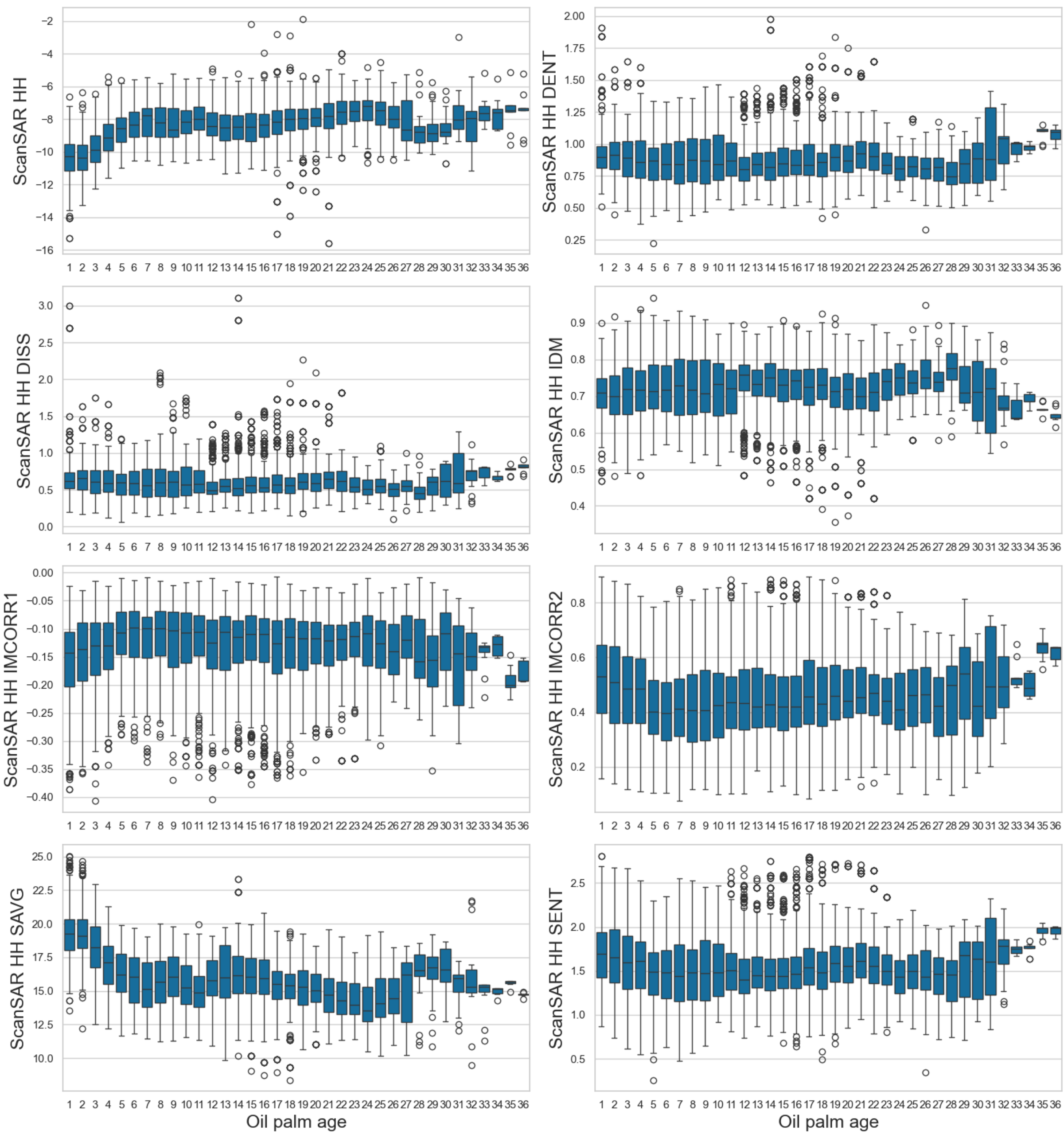

Figure 9. Relationship between PALSAR-2 ScanSAR HH and texture features with oil palm age (absolute value of HH backscattering coefficient considered during texture calculation)

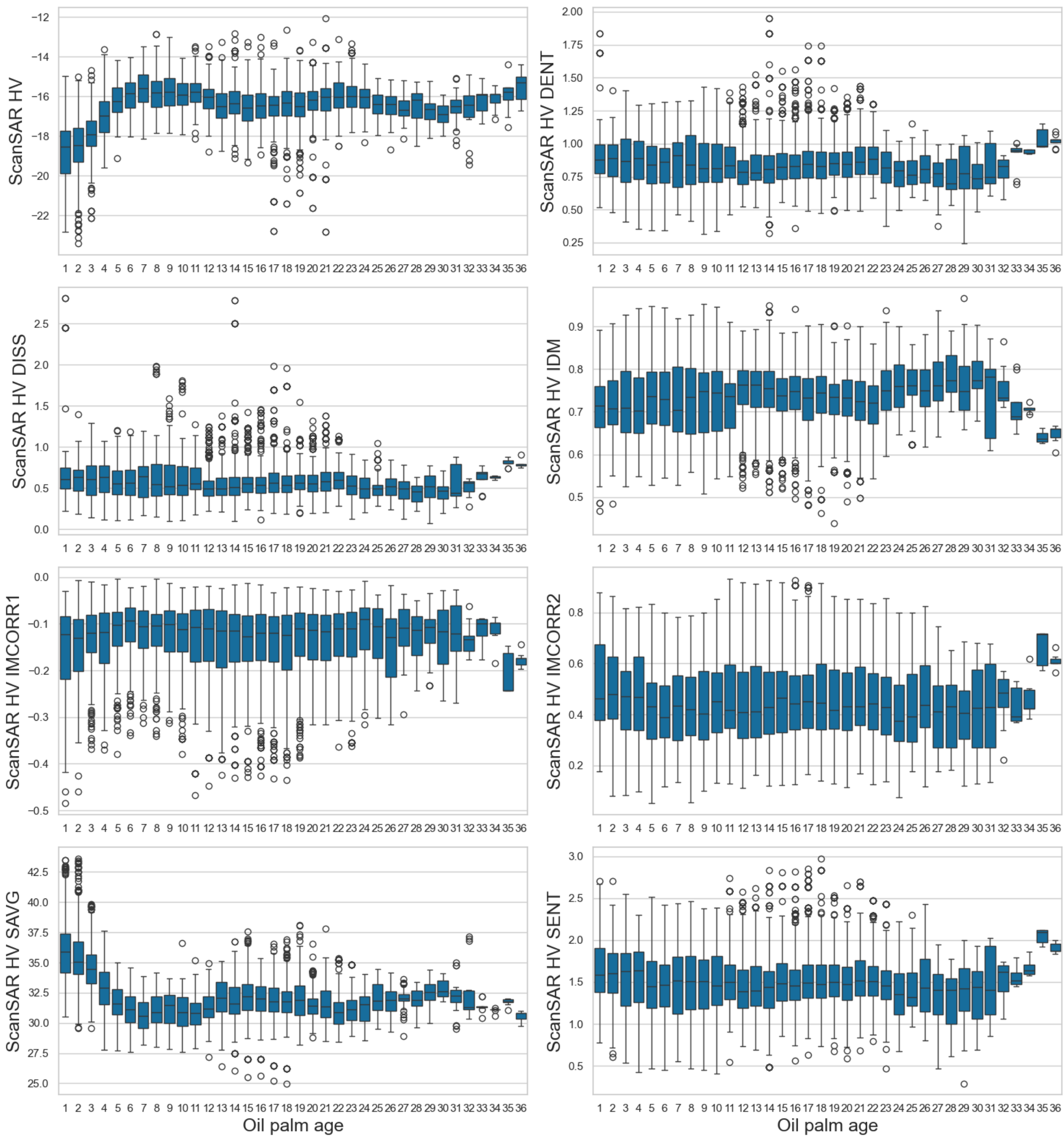

Figure 10. Relationship between PALSAR-2 ScanSAR HV and texture features with oil palm age (absolute value of HV backscattering coefficient considered during texture calculation)

3.3 GEDI forest height

A substantial body of research has established a correlation between the height and age of oil palm trees (Rizeei *et al.*, 2018; Zang *et al.*, 2023). This relationship, however, is subject to many influences, such as the tree's growth environment, agricultural management practices, and genetic predispositions. The growth trajectory of oil palms typically progresses through distinct stages, with the rate and pattern of height increase potentially fluctuating at each stage. A positive correlation exists between the age and height of oil palm trees, suggesting that as the trees age, their height increases proportionally. Nevertheless, this relationship is not strictly linear and may vary based on the growth stages and prevailing conditions (Chong *et al.*, 2017).

In our investigation, we utilized forest height data from the GEDI forest height dataset to examine the relationship between oil palm tree height and age. Oil palms commonly experience rapid growth during the Seed and Young stages, characterized by swift increases in height. The growth rate slows as they enter the Teen stage but remains relatively consistent. In the Mature and Post-mature stages, the height growth rate typically plateaus, transitioning into a slower growth phase. Our findings align perfectly with this growth pattern, further substantiating the reliability of the OPAD.

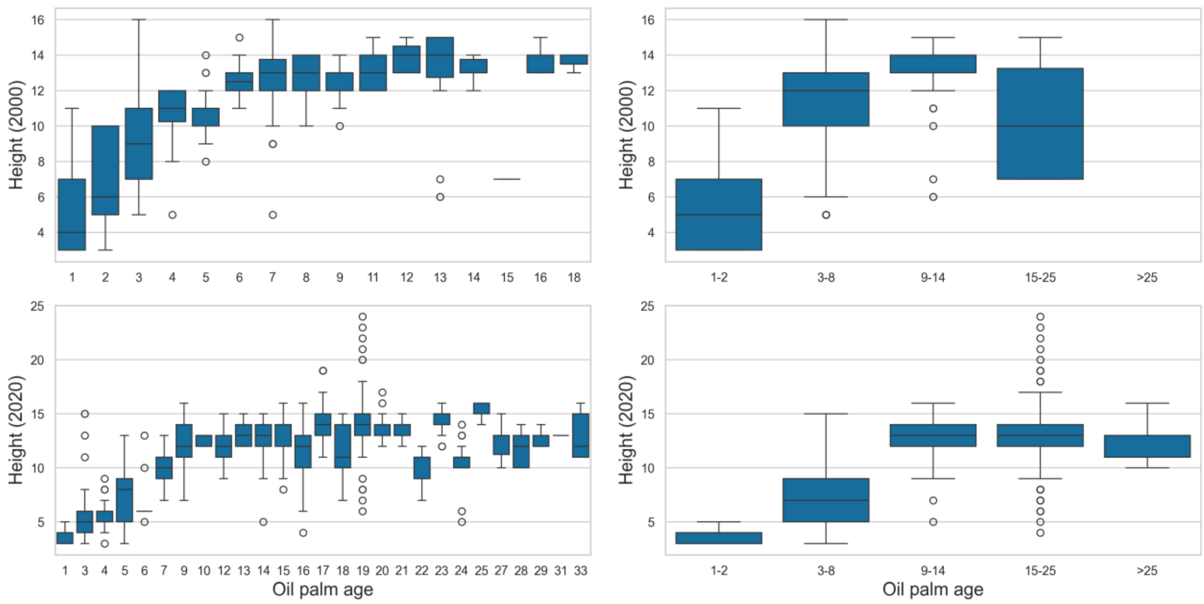

Figure 11. The relationship between oil palm height and oil palm age

4.0 Technical Validation

4.1 Validation accuracy

Table 4 demonstrates that the classification accuracy for oil palm age was enhanced by integrating Planet data sources and GEDI Height data. For the Age category, when validated against Interpreted Samples, the overall accuracy for C2 and C3 improved from 65.56 ± 0.34% to 69.04 ± 0.42% and 77.87 ± 0.70%, respectively. Moreover, the classification accuracy for Field Samples consistently surpassed that of Interpreted Samples, and the Grouped Age category generally outperformed the All Age category. Notably, the C3 achieved the highest validation accuracy when validated with Field Samples for the Grouped Age category, with an impressive overall accuracy of 86.86 ± 0.92%. Even in the All Age category, the validation accuracy for the C3, using Field Samples, reached 84.70 ± 0.99%, a substantial improvement over the C2 under identical conditions. This enhancement is attributed to the robust correlation between oil palm height and age.

Table 4. Validation of the accuracy of OPAD in different scenarios

Category	Interpreted Sample		Field Sample	
	All Age (%)	Grouped Age (%)	All Age (%)	Grouped Age (%)
C1: ScanSAR, Sentinel-1, Landsat-8	65.56 ± 0.34	77.70 ± 0.31	77.14 ± 0.43	83.61 ± 0.38
C2: ScanSAR, Sentinel-1, Landsat-8, Planet	69.04 ± 0.42	79.24 ± 0.29	79.62 ± 0.55	84.45 ± 0.45
C3: ScanSAR, Sentinel-1, Landsat-8, Planet, GEDI forest height	77.87 ± 0.70	83.32 ± 1.11	84.70 ± 0.99	86.86 ± 0.92

Within the Grouped Age category, the C3, when validated against Field Samples, achieved the highest validation accuracy, with an overall accuracy of 86.86 ± 0.92%. Delving into the producer accuracy for specific oil palm age stages, the Young and Mature stages registered the highest accuracies, with 96.08 ± 0.92% and 96.17 ± 1.08%, respectively. Conversely, the Seed, Teen, and Post-mature stages exhibited lower producer accuracies, at 88.57 ± 5.12%, 62.38 ± 1.51%, and 19.09 ± 7.96%, respectively. The post-mature stage recorded the lowest producer accuracy, with many oil palms incorrectly classified as mature. This misclassification may stem from the difficulty of differentiating fully matured post-mature palms from those in the mature stage.

Furthermore, the producer accuracy for the Teen stage was also notably lower, which could be attributed to the rapid growth phase characteristic of this stage, influenced by various factors, including soil conditions, climate, and plantation management practices. This highlights the complexity of accurately predicting the age of oil palms during the Teen stage, as they are frequently misclassified as either Young or Mature.

Table 5. Confusion matrix generated using field samples (grouped age)

Age group	Seed (1-2)	Young (3-8)	Teen (9-14)	Mature (15-25)	Post-mature (>25)	Total	PA (%)
Seed (1-2)	18.6± 1.07	2.4± 1.07	0	0	0	21	88.57 ± 5.12
Young (3-8)	0	49± 0.47	0.1± 0.32	1.9± 0.32	0	51	96.08 ± 0.92
Teen (9-14)	0	0.1± 0.32	13.1± 0.32	7.8± 0.42	0	21	62.38 ± 1.51
Mature (15-25)	0	2.9± 0.88	0.2± 0.42	77.9± 0.88	0	81	96.17 ± 1.08
Post-mature (>25)	0	0.2± 0.42	0	8.7± 0.67	2.1± 0.88	11	19.09 ± 7.96
Total	18.6± 1.07	54.6± 1.43	13.4± 0.52	96.3± 0.95	2.1± 0.88	185	
UA (%)	100.00	89.74 ± 2.21	97.76 ± 3.45	80.89 ± 0.79	100.00		86.86 ± 0.92

4.2 The importance of feature

Evaluating the importance of features is a pivotal step in machine learning and data analysis. This process is instrumental in pinpointing the features that exert the most influence on the prediction of the target variable. It enhances the model's performance and mitigates the risk of overfitting. Moreover, feature importance assessment sheds light on the data's critical elements, bolstering the model's interpretability. In light of this, we assessed the feature importance for Categories 1 to 3 across both the All Age and Grouped Age scenarios. This analysis aimed to ascertain the relative significance of various features and remote sensing data in accurately predicting the oil palm age.

Figure 12 provides a visual representation of the feature importance for Category 3 in both the All Age (Figure 12(a)) and Grouped Age (Figure 12(b)) contexts. The analysis reveals that Planet Blue, GEDI Height, and ScanSAR HH SAVG are paramount for age prediction for the Age scenario. In the Grouped Age scenario, the GEDI Height, ScanSAR HH SAVG, and Planet NIR features stand out as the most influential. However, the Grouped Age scenario accentuates the differences in importance, with most features contributing minimally to the age prediction. The importance is predominantly concentrated on a few key features, with features like IMCORR1, IMCORR2, and SENT showing negligible impact.

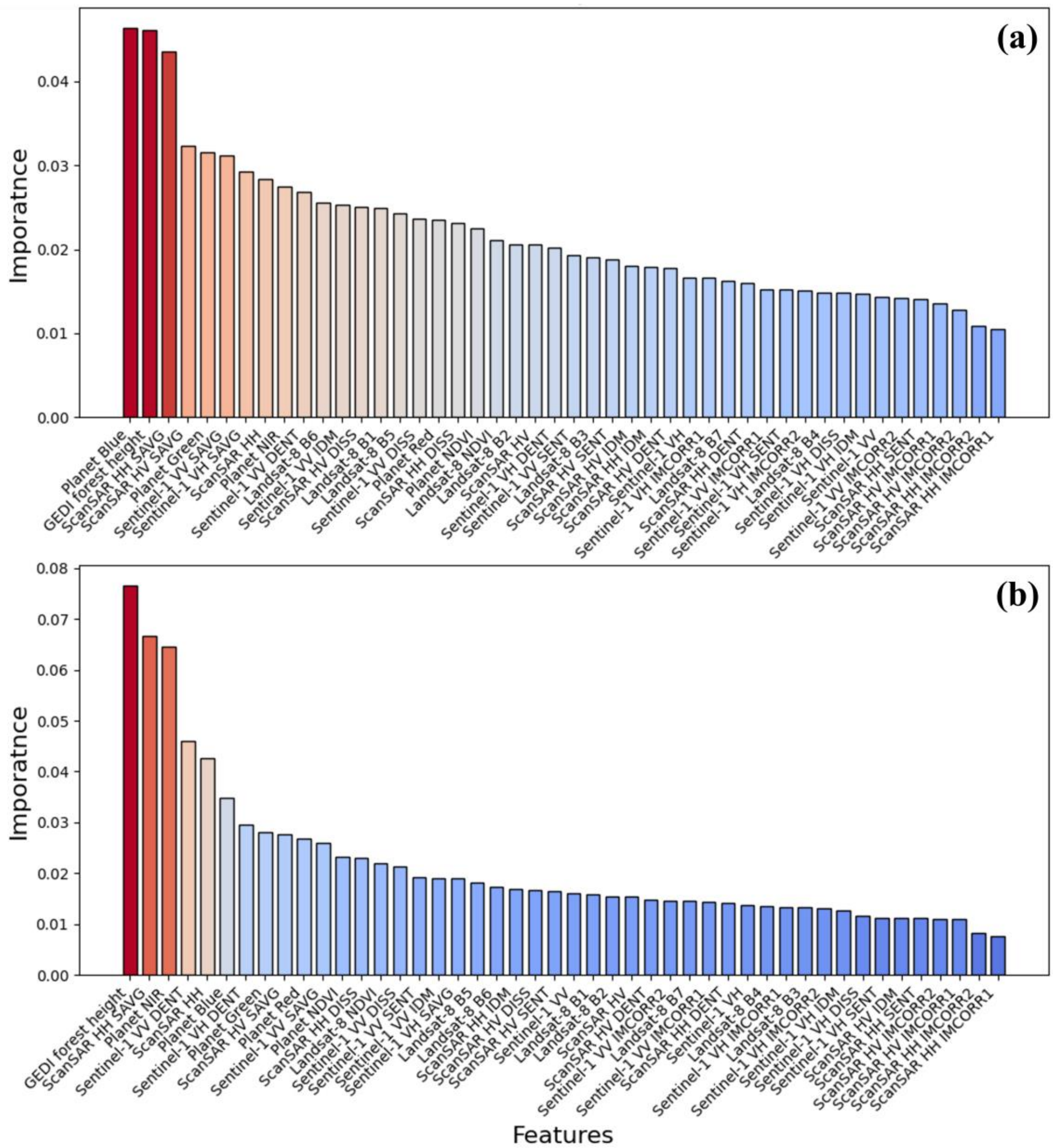

Figure 12. Feature importance ranking for Category 3. (a) under all age scenario; (b) under grouped age scenario

5.0 Discussion and Conclusion

This research effectively illustrates the utility of multisource remote sensing data in accurately predicting the age of oil palms. The development of the OPAD, which is updated annually, offers a wealth of information on the age distribution and associated vegetation traits of oil palms. By amalgamating optical and SAR data, we have examined various remote sensing data, including the NDVI, SAR data texture features, and alterations in GEDI forest height with oil palm age. Notably, spectral features, such as NDVI, Sentinel-1 VV, PALSAR-2 ScanSAR HH, and textural features, such as SAVG, DENT, IDM, and DISS, have shown trends corresponding to the age of oil palms.

The validation outcomes, employing a random forest classifier, underscore the dependability of OPAD for future analyses and its practical application in the sustainable management of oil palm plantations. Features such as GEDI forest height, ScanSAR HH SAVG, Planet NIR, and Planet Blue bands have demonstrated considerable promise in age prediction. These insights are instrumental in refining management practices, evaluating yield potential, and reducing environmental impacts within oil palm cultivation. Future studies could concentrate on broadening the scope of the database, improving the precision of oil palm age data, refining predictive models, and investigating additional factors that influence the growth and productivity of oil palms.

Acknowledgements

This research was supported by the National Key R&D Program of China (2019YFA0606601) and the Tsinghua University Initiative Scientific Research Programs (20223080017, 2021Z11GHX002).

Author Contributions

Qiang Zhao: Writing-Original Draft, Methodology, Investigation, Formal Analysis. **Le Yu:** Conceptualization, Methodology, Validation, Supervision, Writing-Reviewing and Editing. **Zhenrong Du:** Methodology, Validation, Reviewing and Editing. **Kasturi Kanniah:** Resources, Methodology, Reviewing and Editing.

References:

Avtar, R., Ishii, R., Kobayashi, H., Fadaei, H., Suzuki, R., & Herath, S. (2013). Efficiency of Multi-Frequency, Multi-Polarized SAR Data to Monitor Growth Stages of Oil Palm Plants in Sarawak, Malaysia. In 2013 IEEE International Geoscience and Remote Sensing Symposium, 2137-40 Melbourne, VIC, Australia: IEEE.

Camacho, A., Correa, C. V., & Arguello, H. (2019). An analysis of spectral variability in hyperspectral imagery: a case study of stressed oil palm detection in Colombia. International Journal of Remote Sensing, 40(19), 7603-23.

Carolita, I., Darmawan, S., Permana, R., Dirgahayu, D., Wiratmoko, D., Kartika, T., & Arifin, S. (2019). Comparison of optic Landsat-8 and SAR Sentinel-1 in oil palm monitoring, case study: Asahan, North Sumatera, Indonesia. IOP Conference Series: Earth and Environmental Science, 280(1), 012015.

Chemura, A., van Duren, I., & van Leeuwen, L. M. (2015). Determination of the Age of Oil Palm From Crown Projection Area Detected from WorldView-2 Multispectral Remote Sensing Data: The Case of Ejisu-Juaben district, Ghana. ISPRS Journal of Photogrammetry and Remote Sensing, 100(118-27).

Cheng, Y., Yu, L., Cracknell, A. P., & Gong, P. (2016). Oil palm mapping using Landsat and PALSAR: a case study in Malaysia. International Journal of Remote Sensing, 37(22), 5431-42.

Cheng, Y., Yu, L., Xu, Y., Liu, X., Lu, H., Cracknell, A. P., Kanniah, K., & Gong, P. (2018). Towards global oil palm plantation mapping using remote-sensing data. International Journal of Remote Sensing, 39(18), 5891-906.

Chong, K. L., Kanniah, K. D., Pohl, C., & Tan, K. P. (2017). A review of remote sensing applications for oil palm studies. Geo-Spatial Information Science, 20(2), 184-200.

Darmawan, S., Carolita, I., Hernawati, R., Dirgahayu, D., Agustan, Permadi, D. A., Sari, D. K., Suryadini, W., Wiratmoko, D., & Kunto, Y. (2021). The potential scattering model for oil palm phenology based on spaceborne X-, C-, and L-band polarimetric SAR imaging. Journal of Sensors, 2021(1-14).

Fitrianto, A. C., Darmawan, A., Tokimatsu, K., Sufwandika, M., & Iop. (2018). Estimating the age of oil palm trees using remote sensing technique. In *International Conference on Environmental Resources Management in Global Region*.

Hamsa, C. S., Kanniah, K. D., Muharam, F. M., Idris, N. H., Abdullah, Z., & Mohamed, L. (2019). Textural measures for estimating oil palm age. International Journal of Remote Sensing, 40(19), 7516-37.

Li, W., Dong, R., Fu, H., & Yu, L. (2018). Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. Remote Sensing, 11(1), 11.

McMorrow, J. (2001). Linear Regression Modelling for the Estimation of Oil Palm Age from Landsat TM. International Journal of Remote Sensing, 22(12), 2243-64.

Mohd Najib, N. E., & Kanniah, K. D. (2019). Optical and radar remote sensing data for forest cover mapping in Peninsular Malaysia. Singapore Journal of Tropical Geography, 40(2), 272-90.

Mohd Najib, N. E., Kanniah, K. D., Cracknell, A. P., & Yu, L. (2020). Synergy of active and passive remote sensing data for effective mapping of oil palm plantation in Malaysia. Forests, 11(8), 858.

Pohl, C., Kanniah, K. D., & Loong, C. K. (2016). Monitoring Oil Palm Plantations in Malaysia. In 2016 IEEE International Geoscience and Remote Sensing Symposium, 2556-9 Beijing, China: IEEE.

Rizeei, H. M., Shafri, H. Z. M., Mohamoud, M. A., Pradhan, B., & Kalantar, B. (2018). Oil Palm Counting and Age Estimation from WorldView-3 Imagery and LiDAR Data Using an Integrated OBIA Height Model and Regression Analysis. Journal of Sensors, 2018.

Rosenqvist, A., Shimada, M., Suzuki, S., Ohgushi, F., Tadono, T., Watanabe, M., Tsuzuku, K., Watanabe, T., Kamijo, S., & Aoki, E. (2014). Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. Remote Sensing of Environment, 155(3-12).

Shafri, H. Z. M., Ismail, M. H., Razi, M. K. M., Anuar, M. I., & Ahmad, A. R. (2012). Application of LiDAR and optical data for oil palm plantation management in Malaysia. In Lidar Remote Sensing for Environmental Monitoring Xiii, Edited by K. Asai, N. Sugimoto, U. N. Singh, A. Jayaraman, J. Huang and D. Mueller.

Sitorus, J. (2010). Pengembangan Model Estimasi Umur Tanaman Sawit dengan Menggunakan Data Landsat-TM. Jurnal Penginderaan Jauh Dan Pengolahan Data Citra Digital, 1(1).

Tan, K. P., Kanniah, K. D., & Cracknell, A. P. (2013). Use of UK-DMC 2 and ALOS PALSAR for Studying the Age of Oil Palm Trees in Southern Peninsular Malaysia. International Journal of Remote Sensing, 34(20), 7424-46.

Tan, K. P., Kanniah, K. D., & Cracknell, A. P. (2014). On the upstream inputs into the MODIS primary productivity products using biometric data from oil palm plantations. International Journal of Remote Sensing, 35(6), 2215-46.

Toh, C. M., Ewe, H. T., Tey, S. H., & Tay, Y. H. (2019). A Study on Oil Palm Remote Sensing at L-Band With Dense Medium Microwave Backscattering Model. Ieee Transactions on Geoscience and Remote Sensing, 57(10), 8037-47.

Tridawati, A., & Darmawan, S. (2018) Estimation the oil palm age based on optical remote sensing image in Landak Regency, West Kalimantan Indonesia. *IOP Conference Series: Earth and Environmental Science*, 169(1), 012063.

Xu, Y., Yu, L., Li, W., Ciais, P., Cheng, Y., & Gong, P. (2020). Annual oil palm plantation maps in Malaysia and Indonesia from 2001 to 2016. Earth System Science Data, 12(2), 847-67.

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).

Zhao, Q., Yu, L., Li, X., Xu, Y., Du, Z., Kanniah, K., Li, C., Cai, W., Lin, H., & Peng, D. (2024). The expansion and remaining suitable areas of global oil palm plantations. Global Sustainability, 1-20.