JURNAL UPI

Jurnal Geografi Gea

Journal homepage: https://ejournal.upi.edu/index.php/gea

JURNAL GEOGRAFI Gea

Estimating Oil Palm Age Using Regression and Random Forest with Sentinel-2 Data on Google Earth Engine: A Case Study in Betung Krawo, South Sumatra

Anggita Pratiwi Sutrisno[1], Agung Budi Harto[2], Budhy Soeksmanto[3]

[1,2,3]*Geodesy and Geomatics Engineering, Institut Teknologi Bandung, Bandung Indonesia*

Correspondence: E-mail [1]25122003@mahasiswa.itb.ac.id

ABSTRACT

Palm oil (*Elaeis guineensis*) was extensively farmed in Southeast Asia, mainly in Indonesia and Malaysia, and it had a significant impact on the economy of the region. This study aimed to measure and classify oil palm plantations by age using Google Earth Engine, comparing Regression and Random Forest methods. The research focused on Betung Krawo in South Sumatra Province, using Sentinel-2 MSI imagery from 2019 to 2022, with results showing an age range of 1 to 30 years. The Random Forest method achieved an accuracy of 0.844 and a Kappa value of 0.825, and Regression accuracy of 0.922 and a Kappa value of 0.913. Data was divided into 70% for training and 30% for testing in Random Forest Method. For the Regression method, the model derived y= 〖-0.0011x〗 ^2+0.0266x +0.6148, R^2 = 0.8554. which means that the model indicated created fell into the strong category. This research helps to understand the condition and productivity of oil palm plantations, aiding farmers and managers in better decision-making, including early disease detection.

ARTICLE INFO

Article History:
Submitted/Received 09 May 2024
First Revised 12 July 2024
Accepted 30 October 2024
First Available online 31 Oct 2024
Publication Date 31 Oct 2024

Keywords:
Oil palm,
Google Earth Engine,
Random Forest,
Regression,
Age classification

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

1. INTRODUCTION

Oil palm (*Elaeis guineensis*) is a species of palm tree that is widely grown in Southeast Asia, especially in Indonesia and Malaysia. Oil palm is used as an industrial crop that helps in improving the country's economy (Basiron, 2007; Carolita et al., 2015). Indonesia has become the largest CPO exporter in the world by fulfilling ±54% of the world's CPO needs (Sasmito et al., 2019).

In 2019, there was an increase in palm oil production by 12.92%, reaching a total of 48.42 million tons. However, in early 2020, there was a 5.01% decline in CPO production, to 43.41 million tons. In 2021, CPO production decreased again to 45.12 million tons, before finally in 2022 it increased again to 46.82 million tons. For four consecutive years, Riau Province was the largest producer of palm oil in Indonesia (Badan Pusat Statistik, 2019, 2020, 2021, 2022). One of the causes of the significant decline in 2021 in four years was due to the impact of the Covid-19 pandemic (Badan Pusat Statistik, 2021) and also the large number of old plants that were less productive, and were affected by Ganoderma disease or stem base rot (Defitri, 2015; Pramayuda et al., 2021; Hashim, 2018; Kamu, et al., 2015)

This problem, one of the ways to increase palm oil productivity is supported by the President of the Republic of Indonesia who gave Instruction Number 6 of 2019 concerning the "National Action Plan for Sustainable Palm Oil Plantations 2019-2024", one of the actions to increase national palm oil productivity is through strengthening data, strengthening coordination, and infrastructure (Peraturan Menteri Pertanian Republik Indonesia, 2011). However, the problem that is increasingly faced now in the governance of sustainable palm oil plantations is the limited data, information, science, and supporting technology.

Research that has been done (Darmawan, 2016) to increase oil palm productivity is to model the relationship between the age of oil palm growth with ALOS-PALSAR radar satellite images with HH and HV polarization. The results obtained have a correlation value of the model formed is still considered not optimal because the variation in the age of oil palm is only from the growth age of 11 to 21 years. Sample limitation is one of the factors that makes the production estimation model based on ALOS-PALSAR radar satellite imagery produce a model with a low correlation value, $y = 0.8845 \ln(x) - 13.655$; $R^2=0.62$ dan $y=0.6445 \ln(x) - 20.497$; $R^2=0.41$. As for research on the classification of oil palm age using Random Forest, not much has been done. However, there have been several studies conducted related to the classification of oil palms using Random Forest (Tan et al., 2013) with shadow fraction and UK-DMC 2 data to classify the age of oil palms. The method was effective in distinguishing young (3-4 years) and mature (15-29 years) oil palms, improving the overall accuracy by 45.3%. However, the method struggled to distinguish the age of mature trees as the Leaf Area Index (LAI) of oil palms remains stable after about 10 years.

This research aims to map and identify oil palm plantation areas based on age using remote sensing data in cloud computing with Regression and Random Forest methods, especially in the Betung Krawo region in South Sumatra. Prediction of oil palm age is carried out because it can have an important urgency to help related parties such as farmers or companies in managing and optimizing oil palm production, especially in the implementation of replanting or rejuvenation. This data is very useful for early detection of disease if anomalies occur when monitoring oil palm growth. It can also advance a sustainable economy that has significant positive value for farmers, plantation companies, and the environment as a whole (Triyanto, 2019; Verwilghen et al., 2015; Watson-Hernández et al., 2022).

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

2. METHODS

2.1 Study Area

The study area for this research is PT. Plantation Nusantara 7 in Betung Krawo, South Sumatra Province. This oil palm plantation comprises oil palm trees that were planted over a period from 1997 to 2016.

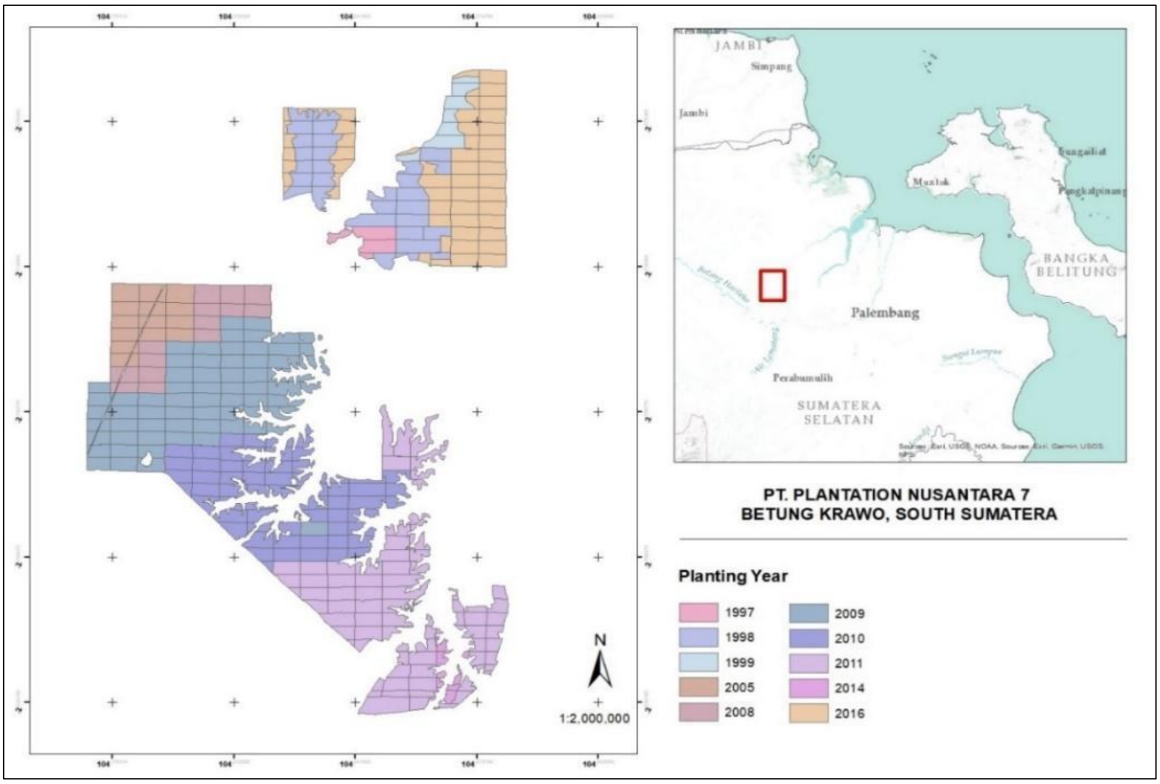


Figure 1. Study Area in PT. Plantation Nusantara 7 Betung Krawo, South Sumatra

2.2 Data

The Level-2A Sentinel-2 MSI data has undergone atmospheric correction, providing more accurate surface reflectance, which supports advanced analyses, such as using vegetation indices like NDVI, EVI, and NDWI (El Hamdi et al, 2024; Fahreza et al, 2022; Ridwana et al, 2022). With 12 different spectral bands, Sentinel-2 MSI offers flexibility for various types of analysis. In this study, data were obtained from the time span between March 1, 2020, and September 30, 2020, provided that the percentage of clouds in each image did not exceed 20%. The following is a list of data used:

Table 1. Research Data

Data	Product	Type Resolution	References
National Palm	LAPAN	Vector	Ministry of Agriculture Directorate of Plantation
Palm Planting Block	LAPAN-PPKS	Vector	(Pramayuda et al., 2021; Darmawan, 2016; Hernawati et al., 2022)
Provincial Administrative Boundary	Geospatial Information Agency	Vector	Geospatial Information Agency
Sentinel-2 MSI: Multi Spectral Instrument Level-2A	ESA Copernicus	Raster 10 meter	ESA Copernicus

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

2.3 Methodology

Figure 2 below depicts the flowchart for the Random Forest and Regression algorithms used in classifying oil palm tree ages. These algorithms are applied to oil palm block areas in three locations: Betung Krawo in South Sumatra.

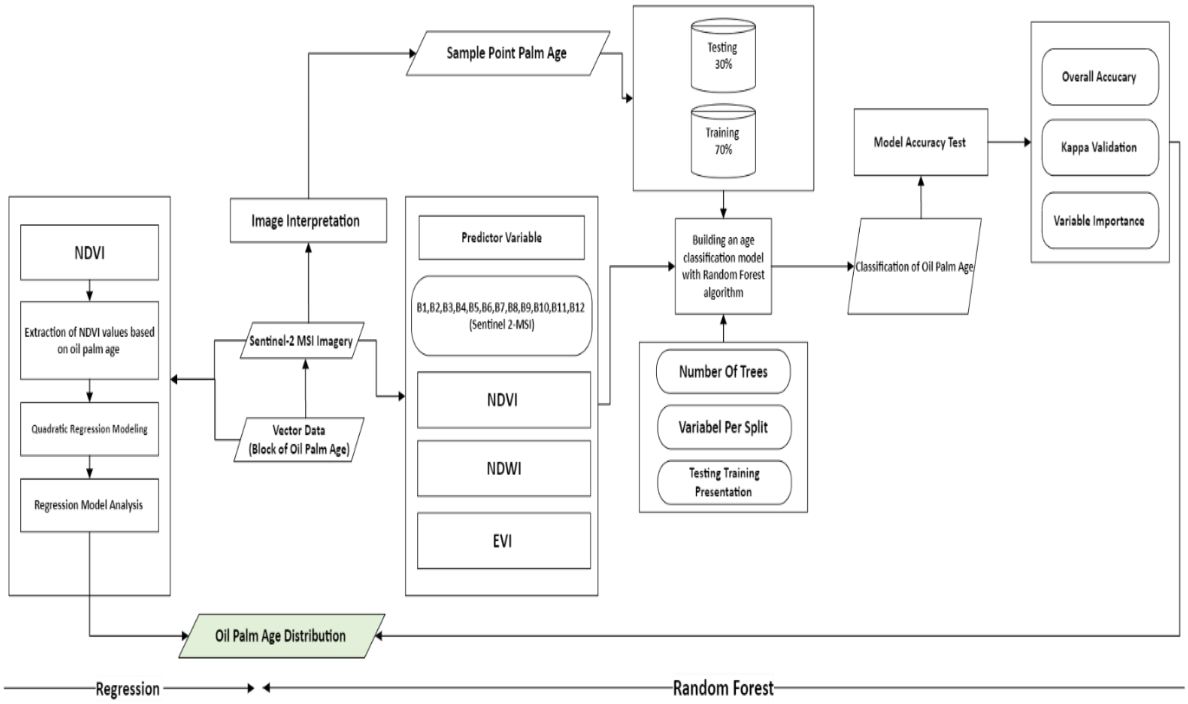


Figure 2. Workflow of Oil Palm Age Classification Using Random Forest and Regression

The data processing consisted of processing using two different methods, however, each produced the same results. Data processing in the regression method starts with data extraction for oil palm age classification, which involves selecting and collecting data from similar sources using Sentinel 2-MSI imagery and observing pixel values. The next step was to create a correlation model using the average NDVI values of the samples sorted by the age of the oil palm blocks.

2.3.1 Regression Method

Step 1: Data Extraction for Oil Palm Age Classification

Data extraction involves selecting and collecting information from several sources that have carried out similar studies (Carolita et al., 2015; Fauzan, 2021; Hernawati et al., 2022; Pramayuda et al., 2021; Tridawati et al., 2018). This extraction process is performed using Google Earth Engine, utilizing Sentinel 2-MSI imagery to analyze pixel values in each dataset. Age sampling is taken from each block of oil palm, with data obtained from LAPAN. The value used in this process is the NDVI (Normalized Difference Vegetation Index) within each planting block.

Step 2: NDVI Data Extraction

This data extraction process is carried out in Google Earth Engine by analyzing each pixel in each data. Data was taken from each age box obtained from LAPAN, taking 30 samples for

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

each age box. The extracted values are NDVI values, which provide information about the health and density of vegetation.

Stage 3: Oil Palm Age Correlation Model

After the NDVI value data extraction process, the extracted sample data was used to create a correlation model. This correlation model was created using the regression method, which aims to understand the relationship between NDVI values and oil palm age. The regression method helps in analyzing how strong and significant the relationship between the variables is.

The table below shows the relationship between NDVI (Normalized Difference Vegetation Index) values and age of oil palm in the regression method. NDVI is an index often used to measure vegetation density and health (Aldiansyah et al., 2021; Pisyam, 2024; Maulana et al., 2022), while the age of the oil palm is measured in months.

In the table, the first column shows the NDVI values, while the second column shows the age of the oil palm in months. From the data, it can be seen that the age of the oil palm ranges from 1 year to 30 years with NDVI values ranging from 0.634 to 0.3.

Table 2. NDVI – Oil Palm Age

NDVI	Age	NDVI	Age
0.634	1	0.776	16
0.664	2	0.762	17
0.692	3	0.744	18
0.716	4	0.724	19
0.738	5	0.700	20
0.756	6	0.673	21
0.771	7	0.644	22
0.784	8	0.612	23
0.794	9	0.576	24
0.800	10	0.538	25
0.803	11	0.496	26
0.804	12	0.453	27
0.802	13	0.404	28
0.796	14	0.354	29
0.787	15	0.300	30

From the data, it can be seen that NDVI values generally tend to increase with the age of the oil palm, indicating an increase in the density and health of the oil palm vegetation as it grows. This is to be expected as oil palms generally grow and experience better vegetative development as they age.

2.3.2 Random Forest Method

Step 1: Splitting Data into Training and Testing Sets

In the first step for palm oil tree age classification using Random Forest, collected data is divided into two groups: training data and testing data. The training data is used to build the classification model (Syafitri et al., 2024), while the testing data is used to evaluate the model's accuracy. There is no strict rule for this split, but a common approach is to allocate 70% of the data for training and 30% for testing. Data is derived from oil palm age blocks using

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

Sentinel 2-MSI imagery, including variables from bands 1 through 12, as well as NDVI, NDWI, and EVI. The goal is to find the proportion that yields the best accuracy for classifying the age of palm oil trees.

Step 2: Bootstrapping Process

Training data is used to build a classification model with the Random Forest algorithm (Syafitri et al., 2024), where patterns in predictor variables are analyzed. Bootstrapping involves drawing random samples from the original dataset to create new datasets, known as Bootstrap Datasets, using "sampling with replacement," which means a sample can be selected more than once. The Bootstrap Dataset has the same size as the original dataset. Samples not selected during bootstrapping form the Out of Bag (OOB) dataset, which is used to measure the model's accuracy. This approach allows Random Forest to assess accuracy internally without a separate testing dataset.

Step 3: Node Splitting with Gini Impurity and Decision Tree Construction

Users can set the number of predictor variables used in each decision tree through the "variables per split" parameter. The Random Forest algorithm uses Gini impurity to select predictor variables and determine the threshold for each split. Gini impurity measures the quality of splits; the lower the Gini value, the more homogeneous the resulting groups. The algorithm seeks combinations of features and thresholds that produce class separations with the lowest Gini impurity, resulting in more efficient decision trees for classification. The process of building decision trees continues until optimal accuracy is achieved, with the number of trees set by the "number of trees" parameter. While more trees tend to increase accuracy, they also require more computation time. Users can also control model complexity with parameters like "Min Leaf Population" and "Max Nodes."

Step 4: Model Deployment and Evaluation

Once optimal accuracy is achieved, the constructed decision trees or model can be applied to the entire study area for classification. Predictor variable values are used as input into the built decision trees to generate output values for the entire area. In the case of classification, the final value for each pixel is determined by a majority vote from the decision trees, resulting in a single value for each pixel. To assess model performance, the testing data set aside during the initial step is used. The predicted outcomes are compared with actual values to evaluate the model's accuracy.

3. RESULTS AND DISCUSSION

3.1 Regression Method

Quadratic regression equations were calculated to analyze data related to palm age and NDVI (Normalized Difference Vegetation Index) values (Carolita, 2019; Fauzan, 2019; Hernawati et al., 2022; McMorrow, 2001; Tridawati et al., 2018; Rizeei, 2018) from Sentinel-2 MSI satellite images. In this study, the mathematical model obtained was +0.0266x +0.6148, R^2 = 0.8554. In this model, (y) is the predicted NDVI (Normalized Difference Vegetation Index), while (x) is the variable used to predict NDVI, which is the age of oil palm.

The negative quadratic coefficient (-0.0011) indicates that there is a negative relationship between the square of oil palm age and NDVI, which means that as oil palm age increases, NDVI values tend to decrease in the long term. On the other hand, the positive linear coefficient (0.0266) indicates that in the early stages of growth, an increase in oil palm age

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

leads to an increase in NDVI values. However, this effect decreases as the age of the oil palm increases due to the negative influence of the quadratic coefficient. This means that as oil palm age (x) increases, NDVI values (y) tend to decrease.

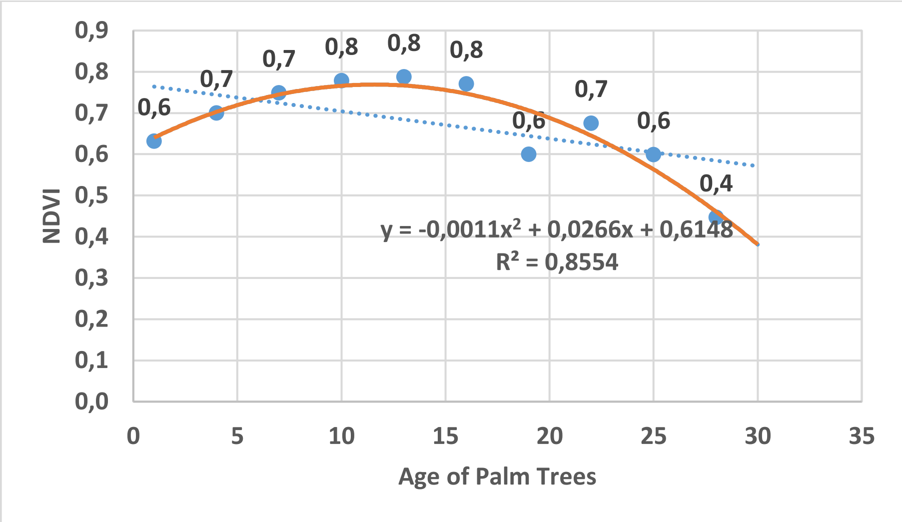
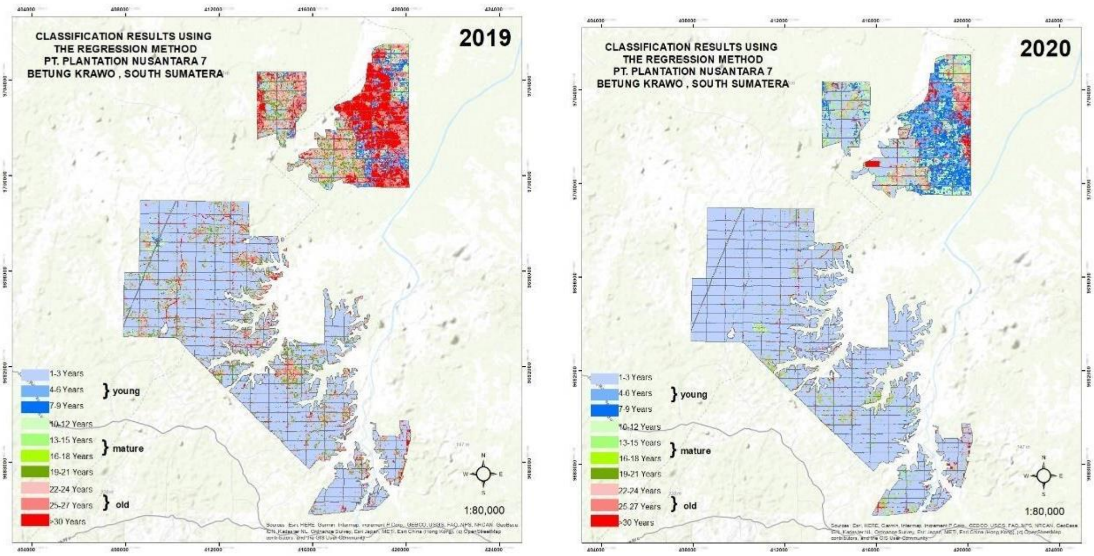


Figure 3. Regression Of Quadratic

In other words, younger palms tend to have higher NDVI values, while older palms tend to have lower NDVI values. The graph shows that NDVI values peak around 8-10 years of age and then start to decrease as the oil palm ages. In this context, the value of $R^2 = 0.8554$ means that about 85.54% of the variability in oil palm age can be explained by the variability in NDVI values. This is an indication that there is a fairly strong relationship between NDVI and oil palm age, and this value is high enough to suggest that NDVI is a significant and relevant predictor of oil palm age. The figure below is a visualization of the age classification of oil palms using the regression method in the Betung Krawo region, Sumatra.



DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

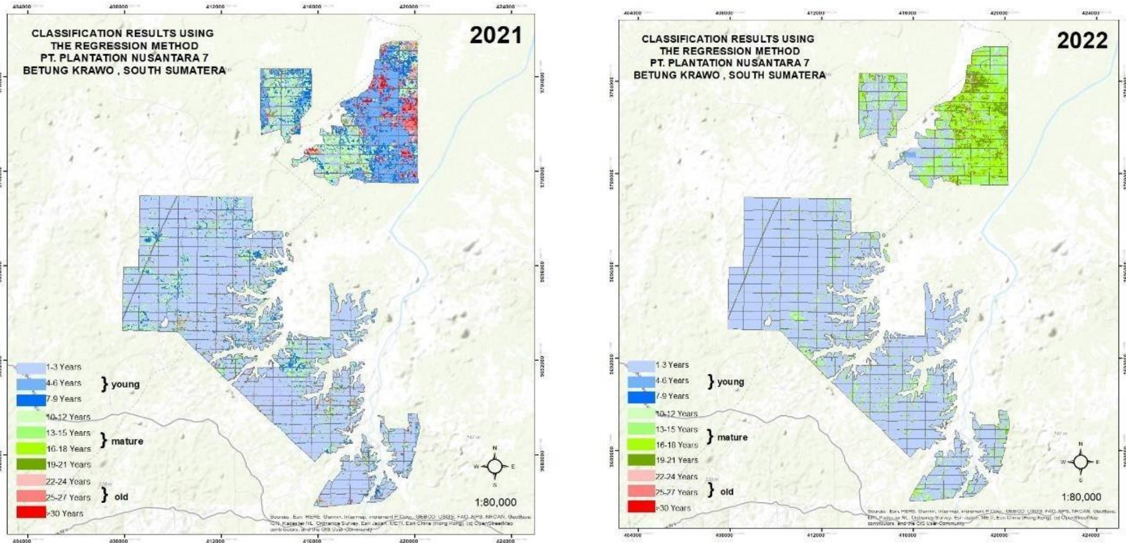


Figure 4. Classification of Oil Palm Age Using Regression Method

The age of oil palm plants within a planting block is often not uniform. Factors such as variety, planting uniformity, cultivation techniques, and different treatments can affect plant growth in different parts of the planting block. However, the age difference within a planting block is usually not very large, around 1-3 years. This age difference is generally caused by phased planting, known as "pulsing", where planting is done in several stages so that at harvest time, the oil palm fruits are of different ages and can extend the fruit production period and maintain production consistency.

In addition, age differences can also occur due to the application of intercropping because oil palm plants must be replanted because the plants are no longer able to produce well or are affected by disease when they are not yet in their productive period. Intercropping is the planting of other crops such as legumes between rows of oil palm with the aim of improving nitrogen content in the soil, maintaining soil structure, maximizing land utilization until the planting area is capable of being planted with oil palm again (Kuvaini et al., 2022).

3.2. Random Forest

In this study, parameter tuning was conducted to try various combinations of parameters to find the combination that gives the best classification results. Results are measured through overall accuracy and validation kappa based on the combination of variables per split and number of trees. Parameter tuning also aims to ensure the model does not experience overfitting or underfitting and can provide the best performance on data that has never been seen before.

Parameter tuning for variables per split was done with 4, while the number of trees used ranged from 100 to 1000, and used a combination of 70% training, 30% testing datasets.

Table 3. Overall Accuracy and Validation Kappa

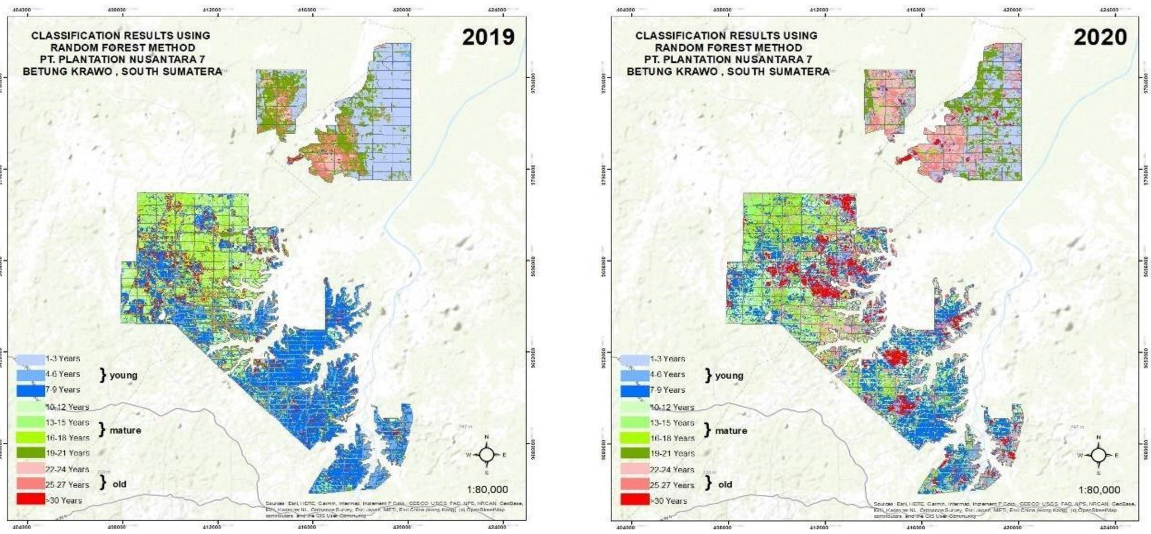
Betung, South Sumatra 70% Testing dan 30% Testing					
Number of Trees	Variable Split	2019	2020	2021	2022

		Overall Accuracy	Validation Kappa	Overall Accuracy	Validation Kappa	Overall Accuracy	Validation Kappa	Overall Accuracy	Validation Kappa
100	4	**0.673**	**0.635**	0.555	0.506	0.536	0.487	0.652	0.614
200	4	0.660	0.621	0.562	0.513	0.524	0.473	0.665	0.628
300	4	0.660	0.621	**0.582**	**0.537**	0.536	0.487	0.652	0.614
400	4	0.673	0.635	0.575	0.529	0.536	0.486	0.652	0.614
500	4	0.667	0.628	0.555	0.506	0.530	0.480	0.658	0.621
600	4	0.654	0.614	0.548	0.498	0.536	0.486	0.652	0.614
700	4	0.654	0.614	0.562	0.514	0.542	0.493	0.646	0.607
800	4	0.642	0.600	0.555	0.506	0.542	0.493	0.652	0.614
900	4	0.648	0.607	0.562	0.514	0.536	0.486	0.646	0.607
1000	4	0.642	0.600	0.555	0.506	0.536	0.486	0.652	0.614

From 2019 to 2022, parameter tuning analysis using Random Forest in Betung, South Sumatra, shows variations in model performance based on the number of trees used in the training process. The table shows that there are fluctuations in evaluation metrics such as Overall Accuracy and Validation Kappa from year to year. In general, there is no consistent trend of increasing or decreasing model performance as the number of trees increases.

For 2019, the model with 100 trees had an Overall Accuracy of 0.673 and Validation Kappa of 0.635, while for 2020, the model with the same number of trees showed lower values with an Overall Accuracy of 0.660 and Validation Kappa of 0.621. In 2021, there is an improvement in performance with some configurations, but there is no clear pattern from year to year.

This suggests that optimization of the Random Forest model in Betung, South Sumatra, requires more in-depth parameter adjustments and possibly the addition of other factors such as tree depth or number of features per split to significantly improve model consistency and accuracy. Figure 5 is the visualization of classification using Random Forest.



DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

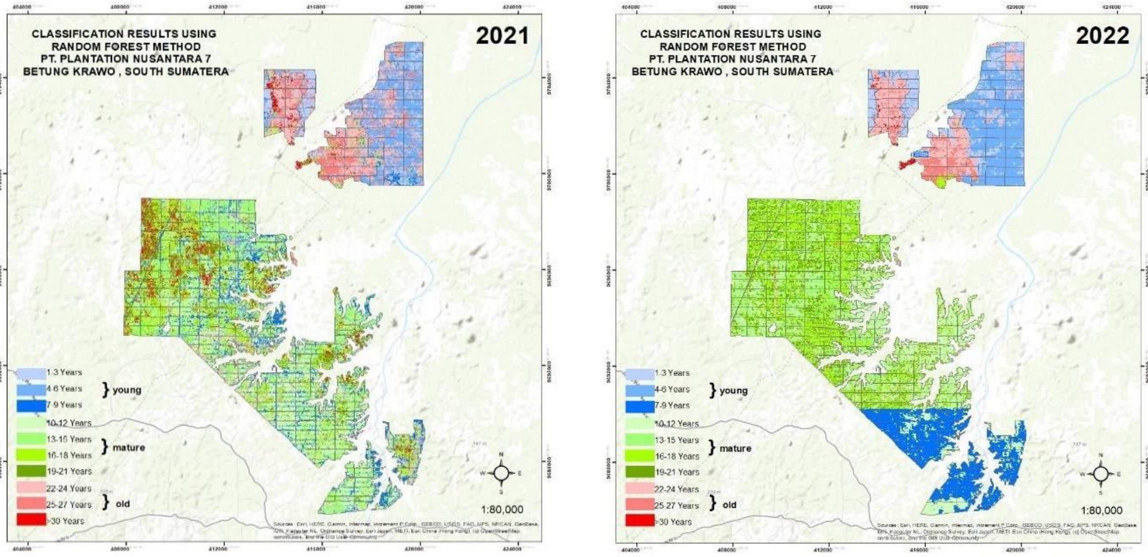


Figure 5. Classification of Oil Palm Age Using Random Forest Method

Errors can still occur when performing classification using Random Forest due to overfitting, especially when applied to very small or overly complex datasets, leading to misclassification of classes. Overfitting can be worsened by a large number of classes, as the model tends to learn intricate patterns that only apply to specific training data. Particularly in cases where the observed area is very small, the variability in the training data can be low and the noise level high, allowing the Random Forest model to fit too closely to this noise instead of capturing the true patterns in the data.

In the context of palm oil tree age classification, some classes might have high correlations among them, especially in very small regions. This can lead the Random Forest model to learn patterns specific only to certain training samples, indicating overfitting. A limited number of training samples can push the Random Forest model to aggressively learn minute details from the data, resulting in the creation of complex decision trees that might not generalize well to unseen data.

3.2.1 Variable Importance

Variable Importance in Random Forest is a measure that indicates how important each input variable is in predicting the response or target variable. In the context of Random Forest, Variable Importance helps identify which features are most influential in the prediction model and optimizes the model by reducing the dimensionality of the data. High values of Variable Importance are considered very important because they contribute significantly, low values of Variable Importance are considered less important because they make little or no contribution to the model prediction.

The difficulty in determining oil palm age from satellite data is that age information cannot be easily extracted due to similar spectral signatures among different oil palm age groups as well as confusion with forest spectral signatures and other features. Therefore, various image processing techniques such as band ratio and vegetation index were used to generate variable image parameters from the original data to overcome spectral confusion among different age groups as well as with other features.

The results show that the existing spectral, multispectral and vegetation index bands help in determining the age of the oil palm. The following is the variable importance of each predictor variable used in the classification process using Random Forest to determine the age of the oil palm:

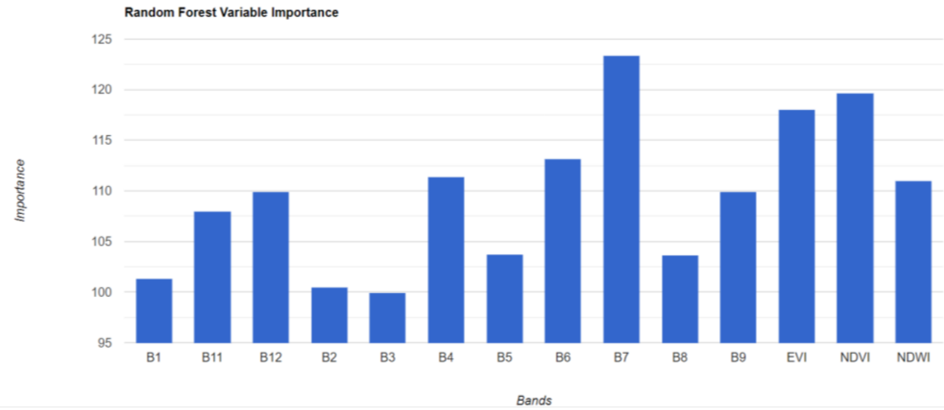


Figure 6. Random Forest Variable Importance

In the classification process using Random Forest to determine the age of oil palm, the importance of predictor variables can be measured through variable importance analysis. The variable importance analysis of the Random Forest (RF) model shows that some predictor variables have a more significant influence in the classification of oil palm age than others. All three graphs show that variables such as NDWI (Normalized Difference Water Index), EVI (Enhanced Vegetation Index), and NDVI (Normalized Difference Vegetation Index) have a high level of importance. This is also in line with the classification results using the regression method, where the NDVI index has a high role in predicting age. This indicates that vegetation condition and soil moisture play an important role in determining the age of oil palms.

B11, B12, B4, B6, B7, and B9 play an important role in palm age classification due to their ability to detect variations in moisture and vegetation conditions. Bands 11 and 6, which are part of the shortwave infrared spectrum (SWIR1), and Bands 12 and 7 of the SWIR2 spectrum, are particularly effective in measuring soil and vegetation moisture. Changes in moisture may indicate differences in the age of the palms, as older trees may show different moisture patterns compared to younger trees. Band 4 (NIR) plays a role in identifying differences in reflectance of healthy vegetation, where older vegetation may show changes in reflectance due to changes in leaf structure and biomass. Band 9 assists in atmospheric correction by detecting water vapor, ensuring more accurate vegetation data is received. The combination of information from these bands allows the Random Forest model to effectively distinguish the age of the palms based on variations in reflectance patterns and moisture conditions.

3.3 Validation

Validation in this study is done by applying the Random Forest (RF) and Regression models to other regions using the *on-site* sampling method or *real* sampling. This means that after the Random Forest model is built using data from one region, the model is tested on another region using sample data collected directly in the field. This *on-site* sampling involves collecting real data at the validation site, which is then used to test the model's ability to classify data in a different region than the one used for model training.

Validation of the Random Forest and Regression models was conducted on oil palm planting blocks at PT Rambang Agro Jaya in Palembang, South Sumatra. This validation was conducted randomly to reflect real field conditions. However, due to flooding in the PT Rambang Agro Jaya area, validation point sampling was only conducted on young and mature

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

oil palms, i.e. from 1 to 20 years old. Other classes could not be included in this sampling. The following are the points from the on-site sampling that were classified as young and mature:

Table 4. Training On-Site

	X	Y	Age (years)
A	490184.2	9623935	1-3
B	488307	9623329	3-8
C	488258.9	9623471	5-10
D	489684.3	9624434	8-15
E	489615.9	9623740	13-20
F	489754.7	9624219	4-8
G	490173.2	9623154	1-8
H	488301.5	9623143	6-10
I	489367.8	9623660	3-6
J	489745.4	9624508	8-15

3.3.1 Validation – Random Forest

The parameter tuning analysis of the Random Forest method for palm oil age classification at PT Rambang Agro Jaya shows that the accuracy and kappa value vary based on the percentage of training and testing data and the number of trees used. In the dataset with 70% training and 30% testing, the highest accuracy was achieved at 200 trees with a value of 0.844, while the highest kappa value was 0.825 at the same number of trees. The following is a visualization and Table of the classification using Random Forest:

Table 5. Tuning Parameter PT. Rambang Agro Jaya (70%,30%) – Random Forest

PT. Rambang Agro Jaya, Palembang 70% Training dan 30% Testing										
Number of Trees	100	200	300	400	500	600	700	800	900	1000
Variabel Per- Split	2	2	2	2	2	2	2	2	2	2
Overall Accuracy	0.811	0.844	0.824	0.824	0.824	0.824	0.824	0.824	0.824	0.824
Validation Kappa	0.789	0.825	0.804	0.804	0.804	0.804	0.804	0.804	0.804	0.804

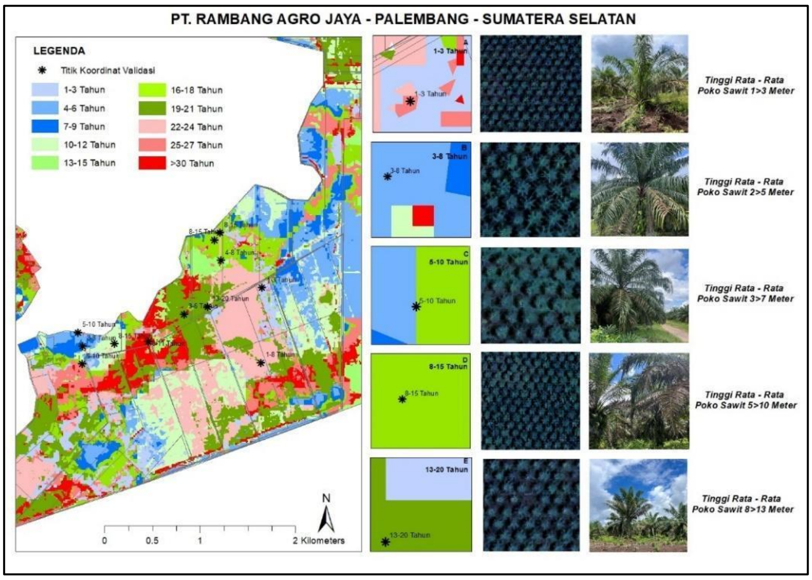


Figure 7. Validation of Random Forest Method (A-E)

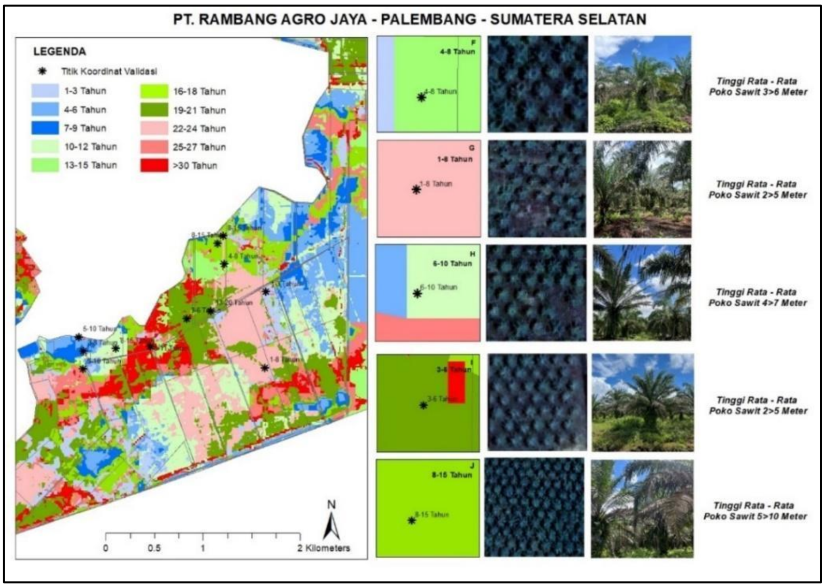


Figure 8. Validation of Random Forest Method (F-J)

3.3.2 Validation – Regression

The following are the Overall Accuracy and Validation Kappa results obtained in the validation process using the Regression method using the 2019 model with the influence of the number of trees, variables per split, and the combination of testing, and training data:

Table 6. Tuning Parameter PT. Rambang Agro Jaya (70%,30%) – Regression

PT. Rambang Agro Jaya, Palembang, South Sumatra 70% Training dan 30% Testing										
Number of Trees	100	200	300	400	500	600	700	800	900	1000
Variabel Per- Split	2	2	2	2	2	2	2	2	2	2
Overall Accuracy	0.915	0.909	0.922	0.909	0.915	0.915	0.915	0.922	0.922	0.915
Validation Kappa	0.905	0.898	0.913	0.898	0.905	0.905	0.905	0.913	0.913	0.905

The parameter tuning analysis of the regression method for oil palm age classification at PT Rambang Agro Jaya showed variations in accuracy and kappa values based on different training and testing datasets. On the dataset with 70% training and 30% testing, the highest overall accuracy reached 0.922 with 800 trees, while the highest validation kappa value was 0.913 on the same number of trees. Small fluctuations in accuracy and kappa were seen with variations in the number of trees, but there was no significant improvement with increasing the number of trees above 800. The validation results that have been carried out at PT. Rambang Agro Jaya, Palembang, South Sumatra Province:

Table 7. Conclusions

	Regression	Random Forest
Overall Accuracy	0.922	0.844
Validation Kappa	0.913	0.825
Number Of Trees	800	200
Variabel Per Split	4	4

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

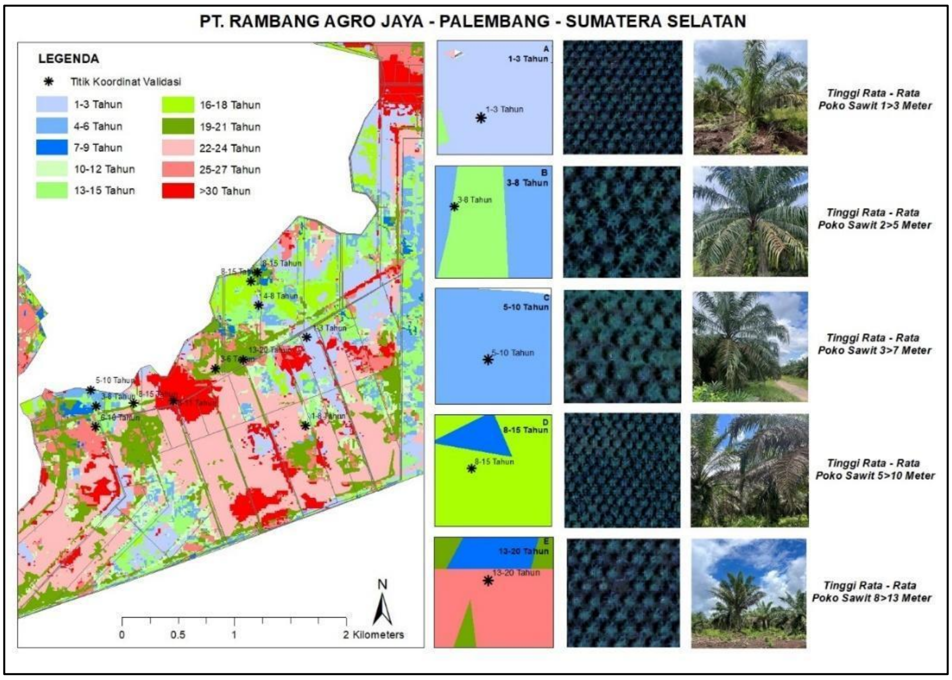


Figure 9. Validation of Regression Method (A-E)

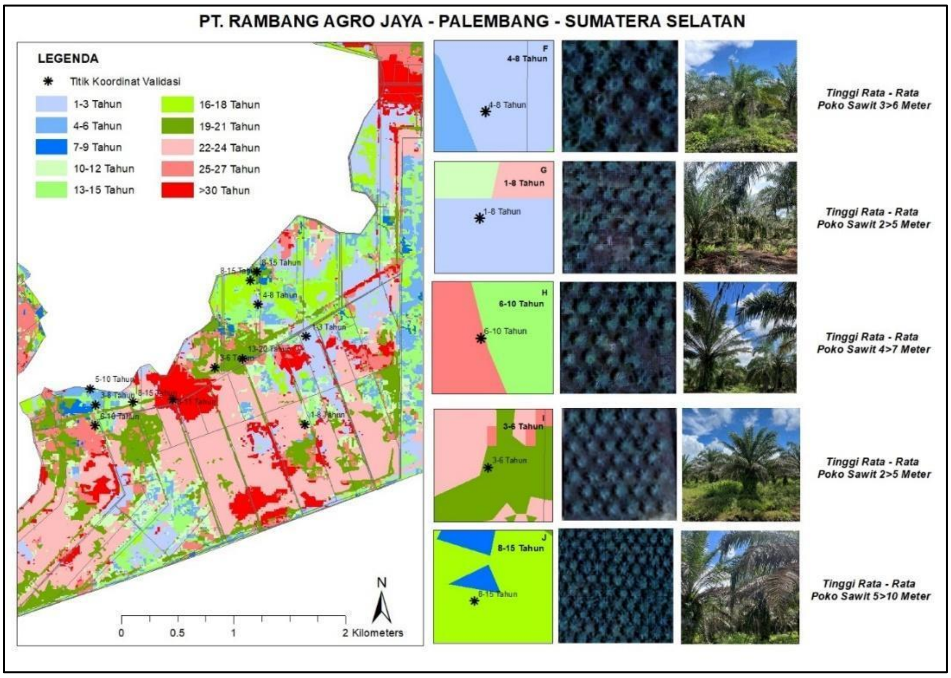


Figure 10. Validation of Regression Method (F-J)

4. CONCLUSIONS

Based on the results of the research and analysis that has been carried out, it can be concluded that the regression method has a better ability than the Random Forest method in predicting the age of oil palm using Sentinel-2 data in the study areas of South Sumatra. It should be noted, that in the validated model, the age of the oil palms depicted only includes young and mature ages, i.e. 1 to 20 years old. This is due to unfavorable weather constraints, i.e. flooding, which occurred during the on-site training sampling for the old oil palm age class.

Sentinel-2 data has also proven to be very effective in observing vegetation, including oil palms. Vegetation indices such as NDVI (Normalized Difference Vegetation Index), EVI

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

(Enhanced Vegetation Index), and NDWI (Normalized Difference Water Index) produced from Sentinel-2 data are very helpful in the palm oil age classification process and consistently become significant predictor variables in predicting palm oil age. This is also supported by the results obtained in the Regression method which states that the NDVI index has a high correlation with oil palm age. These indices provide rich information about the health and development of vegetation, which can be used to distinguish between different ages of oil palms more accurately.

5. RECOMMENDATIONS

To improve the accuracy and applicability of oil palm age classification models, several strategies can be considered. Firstly, using data that covers the entire spectrum of oil palm ages will ensure that the model is not biased towards any particular class and can learn from variations in plant age across the board. Second, utilizing high-resolution imagery will help the model capture better details of vegetation and field conditions, while multispectral or hyperspectral imagery provides in-depth information on plant health. Third, reducing the number of age classes to fewer categories, such as young, mature and old, can simplify the classification problem and improve accuracy by reducing noise and the risk of overfitting. This approach will allow the model to focus on the main growth patterns that are significant for oil palms.

6. REFERENCES

Aldiansyah, S., Mannesa, M. D. M., & Supriatna. (2021). Monitoring of Vegetation Cover Changes in Relationship to Geomorphological Forms Using Google Earth Engine in Kendari City. *Jurnal Geografi Gea*, *21*(2), 159–170.

Badan Pusat Statistik. (2019). *Statistik Kelapa Sawit di Indonesia 2019* (S. S. T. Perkebunan, Ed). Badan Pusat Statistik.

Badan Pusat Statistik. (2020). *Statistik Kelapa Sawit di Indonesia 2020* (S. S. T. Perkebunan, Ed). Badan Pusat Statistik.

Badan Pusat Statistik. (2021). *Statistik Kelapa Sawit di Indonesia 2021* (dan P. Direktorat Statistik Tanaman Pangan, Hortikultura, Ed). Badan Pusat Statistik.

Badan Pusat Statistik. (2022). *Statistik Kelapa Sawit di Indonesia 2022* (S. S. T. Perkebunan, Ed). Badan Pusat Statistik.

Basiron, Y. (2007). Palm oil production through sustainable plantations. *European Journal of Lipid Science and Technology*, *109*(4), 289–295. https://doi.org/10.1002/ejlt.200600223

Carolita, I. (2019). Comparison of Optic Landsat-8 and SAR Sentinel-1 in Oil Palm Monitoring, Case Study: Asahan, North Sumatra, Indonesia. In *IOP Conference Series: Earth and Environmental Science*, 280(1). https://doi.org/10.1088/1755-1315/280/1/012015

Carolita, I., Sitorus, J., & Rizqi, I. S. (2015). Classification of vegetation and oil palm age using SPOT6: The case of tana Laut South Kalimantan, Indonesia. *ACRS 2015 - 36th Asian Conference on Remote Sensing: Fostering Resilient Growth in Asia, Proceedings*.

Darmawan, S. (2016). An investigation of Age and Yield of Fresh Fruit Bunches of Oil Palm Based on ALOS PALSAR 2. In *IOP Conference Series: Earth and Environmental Science*, 37(1). https://doi.org/10.1088/1755-1315/37/1/012037

Defitri, Y. (2015). Identifikasi Patogen Penyebab Penyakit Tanaman Sawit (Elaeis guineensis Jacq.) di Desa Bertam Kecamatan Jambi Luar Kota. *Jurnal Ilmiah Universitas Batanghari Jambi*, *15*(4), 129–133.

El Hamdi, B., Al-Fariziz, M. S., Safitri, M., Amalia, I. N., Nandi, N., & Ridwana, R. (2024). Perbandingan Penggunaan Citra Landsat 8 dan Sentinel 2 Terhadap Analisis Kesehatan Mangrove dengan Metode NDVI Di Teluk Jakarta. *Jurnal Warta Desa (JWD)*, 6(1), 15-24.

Fahreza, F. D., Aulia, A., Fauzan, F. S., Somantri, L., & Ridwana, R. (2022). Pemanfaatan Citra sentinel-2 dengan metode ndvi untuk perubahan kerapatan vegetasi mangrove di kabupaten Indramayu. *Jurnal Pendidikan Geografi Undiksha*, 10(2), 155-165.

Fauzan, I. (2021). Correlation Model of Oil Palm Age Based on Optical RADAR. *42nd Asian Conference on Remote Sensing, ACRS 2021*.

Fauzan, M. I., Fitrianto, A. C., Carolita, I., Gutiérrez-Velez, V. H., Uriarte, M., DeFries, R., Pinedo-Vasquez, M., Fernandes, K., Ceccato, P., Baethgen, W., Padoch, C., Jie, B. X., Zulkifley, M. A., Mohamed, N. A., Perez, R. P. A., Shafry Untoro, M. A., Indriani, L., Wiratmoko, B., Sasongko, A. H., et al. (2019). Estimation of Above Ground Biomass of Oil palm trees by PALSAR. *IOP Conference Series: Earth and Environmental Science*, *2*(1), 21–30. https://doi.org/10.1109/ICSGRC49013.2020.9232547

Hashim, I. C., Shariff, A. R. M., Bejo, S. K., Muharam, F. M., & Ahmad, K. (2018). Classification for Non Infected and Infected Ganoderma boninense of Oil Palm Trees using ALOS PALSAR-2 Backscattering Coefficient. *IOP Conference Series: Earth and Environmental Science*, *169*(1). https://doi.org/10.1088/1755-1315/169/1/012066

Hernawati, R., Wikantika, K., & Darmawan, S. (2022). Modeling of Oil Palm phenology based on remote sensing data: opportunities and challenges. *Journal of Applied Remote Sensing*, *16*(02). https://doi.org/10.1117/1.jrs.16.021501

Kamu, A., Phin, C. K., Seman, I. A., Wan, H. H., & Mun, H. C. (2015). Spatial statistical analysis of basal stem root disease under natural field epidemic of oil palm. *AIP Conference Proceedings*, *1643*, 13–18. https://doi.org/10.1063/1.4907420

Kuvaini, A., Sari, V. I., & Syahputra, D. (2022). Implementasi Model Tumpang Sari Kelapa Sawit dan Semangka di Perkebunan Kelapa Sawit Rakyat Bagan Sinembah Rokan Hilir Riau. *Jurnal Citra Widya Edukasi*, *14*(1), 1-12

Maulana, I., Adifa, S. F., Ummah, E. N. M., Somantri, L., & Ridwana, R. (2022). Analisis perubahan kerapatan vegetasi mangrove di taman nasional Ujung Kulon menggunakan metode NDVI Citra Landsat 8. ULIN: Jurnal Hutan Tropis, 6(2), 172-178.

McMorrow. (2001). Linear regression modelling for the estimation of oil palm age from Landsat TM. *International Journal of Remote Sensing*, *22*(12), 2243–2264. https://doi.org/10.1080/01431160117188

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529

Peraturan Menteri Pertanian Republik Indonesia. (2011). Peraturan Menteri Pertanian Republik Indonesia Pedoman Perkebunan Kelapa Sawit Berkelanjutan Indonesia (Indonesian Sustainable Palm Oil/Ispo) (Permen No. 19 Tahun 2011). *Permentan No 19 Tahun 2011 Tentang Pedoman Perkebunan Kelapa Sawit Berkelanjutan Indonesia*, 243–248.

Pisyam, A. N. M. (2024). Analysis of Changes in Distribution of Urban Heat Island in Medan City and Its Relationship with Land Change Using Remote Sensing Data. *Jurnal Geografi Gea*, *24*(1), 64–74.

Pramayuda, R., Supriatna, Manessa, M. D. M., & Carolita, I. (2021). Pemetaan Umur Tanaman Kelapa Sawit dengan Pendekatan Machine Learning Menggunakan Google Earth Engine (Studi Kasus : Provinsi Riau). *Prosiding Seminar Geomatika VI*, *August*, 381–390.

Ridwana, R., Al Kautsar, A., Saleh, F., Himayah, S., Arrasyid, R., & Pamungkas, T. D. (2022, November). Spatiotemporal monitoring of rice crops in the COVID-19 pandemic period for local food security using sentinel 2b imagery case study: Tasikmalaya city. In IOP Conference Series: Earth and Environmental Science, 1089(1), p. 012039. IOP Publishing.

Rizeei, H. M. (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*. https://doi.org/10.1155/2018/2536327

Sasmito, Ginanjar S., Laut, L.T., Destiningsih, R. (2019). Daya Saing Crude Palm Oil (CPO) Indonesia dan Malaysia di Limapasar Utama Tahun 2001 – 2018. *Dinamic: Directory Journal of Economic*, *1(*3), 257-268. doi:10.31002/dinamic.v1i3.790.

Syafitri, D. R., Arysandi, S. A., Purboyo, A. A., Fitri, A., Febriani, Y., Tjahyanityasa, W. F., Sari, N. P., Mar'ath, S. K., Safinatunnajah, Satyaningrum, A. C., Rajardjo, N., Sudrajat, Mardiatno, D., & Rachmawati, R. (2024). Flood Assessment Through Integrated-Mixed-Method in The Western Downstream Area of Citanduy River Basin, Pangandaran Regency. *Jurnal Geografi Gea*, *24*(1), 25–48.

Tan, K. (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), 7424–7446. https://doi.org/10.1080/01431161.2013.822601

Tridawati, A., Darmawan, S., & Armijon, A. (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). https://doi.org/10.1088/1755-1315/169/1/012063

Triyanto, Y. (2019). Palm oil prediction production using Extreme Learning Machine. *International Journal of Scientific and Technology Research*, *8*(8), 1070–1072.

Verwilghen, A., Rabillard, M.-A., Chaval, Y., Rieffel, D., Sinaga, M. H., Naim, M., Caliman, J.-P., Giraudoux, P., & Raoul, F. (2015). Relative age determination of Rattus tiomanicus using allometric measurements. *Mammalia*, *79*(1), 81–90. https://doi.org/10.1515/mammalia-2013-0113

Watson-Hernández, F., Gómez-Calderón, N., & da Silva, R. P. (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

DOI: https://doi.org/10.17509/gea.v24i2.69334
p-ISSN 1412-0313 e- ISSN 2549-7529