Discriminating the Severity of Basal Stem Rot Disease in Oil Palm (Elaeis guineensis Jacq.) Plantation Using Sentinel-2

R D Handrian¹, B H Trisasongko¹², and D R Panuju¹²
¹Department of Soil Science and Land Resource, Bogor Agricultural University, Jalan Meranti, Kampus IPB Dramaga, Bogor 16680, Indonesia.
²GITIIA, P4W/CRESTPENT, Bogor Agricultural University, Jalan Pajajaran, Bogor, Indonesia
Email: ramadhani_handrian@apps.ipb.ac.id

Abstract. Oil palm remains the most prominent plantation sector in Indonesia. Monitoring large scale plantations require specificity in datasets and approaches. Observing the extent of plantations has long been conducted using remotely-sensed data. With the complex issues in oil palm plantations, focus has also been made in investigating the utility of earth-observing satellites to estimate foliar nutrients and to identify the impact of plant diseases. The latter has been one of emerging research interests, especially the basal stem rot (BSR) caused by Ganoderma boninense. In this research, we proposed a freely-available, multispectral data of Sentinel-2 taking the advantage of frequent observation period. Preliminary discrimination of BSR impacts was examined by using a popular machine learning approach, i.e. random forests. Ground data as the reference were collected in Cikasungka plantation, Bogor, Indonesia. Using a tuned random forest model, we obtained overall accuracy about 82% with five distinctive targets, i.e. four severity levels and the normal trees. This research suggests that tuned random forest model could be invaluable for constructing a proper machine learning model that adaptive to data feeds.

1. Introduction

Oil palm (Elaeis guineensis Jacq.) plantations produce crude palm oil (CPO), which is often used as cooking oil, industrial oil, and fuel. Oil palm tree is originated from West Africa and has become popular after the industrial revolution leading to a high demand of palm oil. Oil palm is the most prominent plantation sector in South East Asia, especially Malaysia and Indonesia. In the upcoming years, it is estimated that the largest investment in the plantation sector remains in oil palm.

In Indonesia, oil palm plantations are generally managed by state-owned or private corporations, as well as small-scale farmers. Managing large plantations is a challenge in terms of monitoring schemes, therefore, the use of remote sensing technology is an efficient and effective strategy to collect plantation-related data. Plantation sector has increasingly utilized remote sensing technology; details are presented elsewhere, for instance [1–3]. This also applies to oil palm plantations.

Investigations using numerous multispectral sensors with varying spatial resolutions in the tropics have been carried out. Within the context of oil palm plantations, these research include estimating stand age [4], detection of oil palm plantation areas and their dynamics [2,5], yield estimation [6], nutritional deficiencies [7], and monitoring of plant health [8]. This suggests that, over this time span, research foci have become more extensive; hence, it is necessary for earth observers to further assess the utility of
Various studies have presented potential uses of imaging technologies/techniques for detecting disease in plants [9–11]. Pathogens have a detrimental impact on plants in stomatal conductance, photosynthetic pigments, and photosynthetic rates. In addition, noxious compounds in pathogens could cause damage to plants in a way that modifies reflectance of plant leaves [12,13]. Another study using QuickBird imagery to identify diseases and their spatial patterns showed that very high resolution images could be optimized to estimate location and the extent of oil palm infection [14]. Survey of the literature indicates that one of prominent problems in oil palm plantations is related to the impact of basal stem rot (BSR) disease.

BSR disease is caused by *Ganoderma boninense* fungus. This disease has been in concern, causing significant tree mortality in several Indonesian plantations. *G. boninense* degrades lignin into carbon dioxide, water, and cellulose. BSR can easily be identified in the field. Common visual properties include unopened spear leaves, yellowing and necrotic leaves and small canopy coverage. In an expanded stage, fungal fruiting bodies and fallen stems can be observed. In current practice, the identification of the symptoms is based on manual observation, which is expensive and time consuming. With this reason, an alternative approach to detect BSR disease should be developed, based on visual symptoms, which could later be translated into the severity of BSR in a spatial unit. Discrimination of BSR severity can be carried out using machine learning models, including Random Forest (RF).

Studies have shown that RF algorithms can produce a better result [15–17]. RF has a little pre-processing requirement, because it is insensitive to varying data units and can handle unbalanced data [18] Using very high resolution QuickBird imagery for detecting BSR disease, RF was the best at predicting, classifying, and mapping the disease in comparison to Support Vector Machine (SVM) and classification and regression tree (CART) [19]. Nonetheless, research employing multispectral images mostly use images that are inaccessible for free and, some of them, require a long revisit time, leading to a weaker implementation. Freely available remote sensing datasets, including Sentinel-2, are therefore more suited for frequent observation. Unfortunately, the sensor has yet to be fully exploited for BSR detection and mapping.

With this reason, a research on the identification of BSR disease and the investigation of its severity levels using Sentinel-2 is required to assist monitoring over oil palm plantations. This research, therefore, aimed (1) to map the distribution of affected oil palm through a machine learning approach and (2) to investigate the role of vegetation indices available to Sentinel-2 data in discriminating the severity levels.

2. Methodology

2.1. Research Area

This research was situated at Cikasungka plantation of PT. Perkebunan Nusantara VIII, Bogor Regency, specifically at 106°31'10" - 106°31'20" E and 6°32'10" - 6°32'40" S. The area comprised of Afdeling 3 on Blocks 54 and 66. Block 54 had 17-year old oil palms occupying 26.73 ha, while Block 66 hosted 19-year-old trees covering 26.60 ha (Figure 1).
2.2. Data Collection and Analysis

General processing step is shown in Figure 2. It starts from field observation, pre-processing Sentinel-2, extracting pixel values according to ground sample locations, RF classification in R software, and producing distribution maps of BSR severity disease.

Firstly, coordinates of sample trees were taken utilizing a handheld navigation system, both for healthy and affected trees. Labelling severity was done in situ, based on four different symptoms (Table 1) [20]. In total, there were 85 samples taken during field surveys.

In this research, we used Sentinel-2A multispectral imagery taken on 23 February 2021. This image was downloaded from Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/#/home). It was preprocessed using Sentinel Application Platform (SNAP) by applying subset and resample procedures. Preprocessing was straightforward since level 2A of Sentinel-2 data is bottom-of-atmosphere reflectance. In order to investigate the statistics, pixel values were extracted from the imagery. Extracted pixel values were determined by rounded polygon (buffer) around field samples, performed using QGIS 3.16 software. This research used 7.5 m diameter from sample coordinates, considering planting space. To ease the analysis, only bands with spatial resolutions of 20 m and 10 m were considered (Table 2). With this setting, band 1 (coastal aerosol), band 9 (water vapor) and band 10 (SWIR – Cirrus) having 60 m spatial resolution were neglected. All selected bands were resampled onto 10 m for further analysis.

Vegetation indices were retrieved to understand their role in discriminating the severity levels of BSR. The values are generated by combining data from multiple spectral bands into single image. In this research, evaluated vegetation indices included Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Green Blue Normalized Difference Vegetation Index (GNDVI), Green Blue Normalized Difference Vegetation Index (GBNDVI), Atmospherically Resistant Vegetation Index (ARVI), Soil Adjusted Vegetation Index (SAVI), and Simple Ratio (SR). Details are given in Table 3.

Table 1. Classification of the severity of BSR disease

| Severity        | Damage       | Symptoms                                      | Level | Samples |
|-----------------|--------------|-----------------------------------------------|-------|---------|
| Healthy         | ≤ 5%         | Healthy plants                                | 0     | 22      |
| Light           | >5% - 25%    | Midrib is broken and hanging                  | 1     | 15      |
| Medium          | >25% - 50%   | Spear leaves appear more than 3               | 2     | 15      |
| Critical        | >50% - 75%   | Fruiting body of Ganoderma fungus appears on the stem | 3     | 13      |
| Very critical/death | >75% - 100% | Trees fall and die                            | 4     | 20      |
Figure 2. Research flow

Table 2. Sentinel-2A spectral band used in this research

| Sentinel 2A-Band | Wavelength (nm) | Wide band (nm) | Spatial resolution (m) |
|------------------|-----------------|----------------|-----------------------|
| Band 2 (Blue)    | 492.4           | 66             | 10                    |
| Band 3 (Green)   | 559.8           | 36             | 10                    |
| Band 4 (Red)     | 664.6           | 31             | 10                    |
| Band 5 (Vegetation Red Edge) | 704.1 | 15 | 20 |
| Band 6 (Vegetation Red Edge) | 740.5 | 15 | 20 |
| Band 7 (Vegetation Red Edge) | 782.8 | 20 | 20 |
| Band 8 (NIR)     | 832.8           | 106            | 10                    |
| Band 8A (Narrow NIR) | 864.7 | 21 | 20 |
| Band 11 (SWIR)   | 1613.7          | 91             | 20                    |
| Band 12 (SWIR)   | 2202.4          | 175            | 20                    |

Table 3. Vegetation indices of Sentinel-2A

| Indices | Band Math Formula | Reference |
|---------|-------------------|-----------|
| NDVI    | (B8–B4)/(B8+B4)   | [21]      |
| GNDVI   | (B8–B3)/(B8+B3)   | [22]      |
| GBNDVI  | (B8−(B3+B2)/(B8+B3+B2) | [23] |
| ARVI    | (B8−rb)/(B8+rb); rb= (B4)-gamma*(B2−B4), gamma=1 | [24] |
| SAVI    | (1+L)∗(B8−B4)/(B8+B4+L), L is a canopy background adjustment factor (1 for low, 0,5 for medium and 0 for high density vegetation) | [25] |
| SR      | B8/B4             | [26]      |
In this study, image analysis and prediction employed RStudio software. Pixel values obtained from 10 bands of Sentinel-2 imagery were utilized as variables in modeling. Modeling using machine learning involves data ingestion and partitioning, statistical modeling, predicting and measuring accuracy. Whole samples were separated randomly considering the amount of data contained in each severity class. Data separation into training and testing data was implemented using a ratio of 75% for training and 25% for testing data [27]. Eighty five polygons were rasterized into 153 pixels; hence, whole data were split into 114 and 39 pixels for training and testing, respectively (Table 4).

| Data | Percentage (%) | Level 0 | Level 1 | Level 2 | Level 3 | Level 4 | Total |
|------|----------------|---------|---------|---------|---------|---------|-------|
| Training | 75 | 29 | 23 | 18 | 17 | 27 | 114 |
| Testing | 25 | 10 | 7 | 7 | 5 | 10 | 39 |
| Total | 100 | 39 | 30 | 25 | 22 | 37 | 153 |

This research employed random forest (RF) learning algorithm for classification [28]. Specifically, randomForest R package was chosen for RF classifier, allowing number of trees (ntree) parameter to be tuned. Confusion matrix was used for evaluating models. The last step was producing distribution of severe trees infected with G. Boninense based on the best RF model.

3. Result and Discussion

3.1. Accuracy Assessment

Table 5 shows confusion matrix which represents accuracy of a model using default setting. Overall accuracy was about 82%. Table 5 explains that severe attack (level 4) tended to have higher accuracy, while the rests showed variably, ranging from 70% to 86%. Significant difference of level 4 was due to clear difference as this level has its tree canopy absent (Figure 3). This suggested that distinctive reflectance values were clearly observed and RF model was able to classify the target well. The Jeffries-Matusita distance indicated comparable outcome to this research, especially between healthy trees and light severity trees [29]. Oil palm canopy was found correlated with some spectral bands. Leaf pigment and canopy reflectance indicate that chlorophyll contents serve as a good indicator of photosynthesis and plant health conditions [30].

| Actual | Predicted |
|--------|-----------|
|        | Level 0 | Level 1 | Level 2 | Level 3 | Level 4 | All |
| Level 0 | 7 | 0 | 0 | 1 | 0 | 8 |
| Level 1 | 2 | 6 | 1 | 0 | 0 | 9 |
| Level 2 | 1 | 0 | 5 | 0 | 0 | 6 |
| Level 3 | 0 | 1 | 0 | 4 | 0 | 5 |
| Level 4 | 0 | 0 | 1 | 0 | 0 | 11 |
| All | 10 | 7 | 7 | 5 | 10 | 39 |

Overall’s accuracy (%) 82
Machine learning model may not be optimized in various conditions. Tuning parameters of machine learning models is therefore suggested to explore suitable parameter in order to understand their potential utilization [2,16,17]. While \textit{n}tree parameter is generally pre-determined with 500, this setting may not be optimum. Although the acquired data were fairly imbalance, this research indicated that RF model could adapt, similar to the finding of previous research [31]. Implementing the model into the whole Sentinel-2 image was done using raster package. Figure 4 shows the distribution of reflectance values based on the classification of BSR disease severity. In general, these two oil palm blocks have partially impacted by Ganoderma.
3.2. Sentinel-2 Vegetation Indices
Visible and NIR bands of Sentinel 2 were ingested into band math formula (Table 3) to evaluate their role in discriminating the severity. Theoretically, infrared bands can strongly be reflected by chlorophyll; hence, they are suitable to assess healthy and unhealthy plants. Figure 5 depicts the behaviour of vegetation indices according to severity levels.

NDVI makes use of the physical phenomenon of reflected light waves from leaves, scaling the responses into a range of -1 to 1. In general, high NDVI values were found in healthy stands (Figure 5). Low impacted trees were found indistinguishable to healthy trees with this measure. This might be due to existence of green leaves in both levels 1 and 2. In devastated trees, chlorophyll was limited; hence, NDVI values were reduced. The patterns were found similar to all vegetation indices, except GNDVI. Due to limited datasets and the extent of GNDVI assessment, the reason is yet to be fully understood. It was suspected that this was due to the use of green band in GNDVI, instead of the red band. This warrants a future research, exploring deviant of the pattern. In general, this research found that contributions of vegetation indices were generally limited. With the possible diminishing effect in RF [32], adding vegetation indices into a set of spectral bands should carefully be taken.

**Figure 5.** Distribution of reflectance values of vegetation indices: (a) NDVI, (b) GNDVI, (c) SR, (d) GBNDVI, (e) SAVI, and (f) ARVI.
4. Conclusion
This research showed that the use of RF model on the Sentinel-2 yielded an overall accuracy of about 82%. This indicates that RF can be invaluable for machine learning modeling for distinguishing the disturbance in oil palm plantation. With readily available dataset, Sentinel-2 has been invaluable to support a better management practice in oil palm plantations. Additional datasets like vegetation indices showed insensitive to the discrimination, although they helped to distinguish severely impacted oil palm trees due to Ganoderma.

Acknowledgments
The authors thanks to the management of PT. Perkebunan Nusantara VIII for their permission to carry out this research in the Cikasungka plantation.

References
[1] Trisasongko B H and Paull D J 2020 A review of remote sensing applications in tropical forestry with a particular emphasis in the plantation sector Geocarto Int. 35 317–39
[2] Panuju D R, Paull D J and Trisasongko B H 2019 Combining binary and post-classification change analysis of augmented ALOS backscatter for identifying subtle land cover changes Remote Sens. 11
[3] Trisasongko B H and Panuju D R 2015 Characteristics of L-band backscatter coefficients of rubber plantation and their seasonal dynamics AIP Conf. Proc. 1677 60006
[4] McMorrow J M 1995 Relation of oil palm spectral response to stand age Int. J. Remote Sens. 16 3203–9
[5] Panuju D R and Trisasongko B H 2012 Seasonal Pattern of Vegetative Cover from NDVI Time-Series Trop. For.
[6] Balasundram S, Memarian H and Khosla R 2013 Estimating oil palm yields using vegetation indices derived from QuickBird Life Sci. J. 10 851–60
[7] Amirruddin A D, Muharam F M, Paing T N, Singh D S K and Yusoff M M 2017 Nitrogen effects on growth and spectral characteristics of immature and mature oil palms Asian J. Plant Sci. 16 200–10
[8] Khosrokhani M, Khairunniza-Bejo S and Pradhan B 2018 Geospatial technologies for detection and monitoring of Ganoderma basal stem rot infection in oil palm plantations: a review on sensors and techniques Geocarto Int. 33 260–76
[9] Golhani K, Balasundram S K, Vadamalai G and Pradhan B 2019 Selection of a spectral index for detection of orange spotting disease in oil palm (Elaeis guineensis Jacq.) using red edge and neural network techniques J. Indian Soc. Remote Sens. 47 639–46
[10] Husin N A, Khairunniza-Bejo S, Abdullah A F, Kassim M S M, Ahmad D and Aziz M H A 2020 Classification of basal stem rot disease in oil palm plantations using terrestrial laser scanning data and machine learning Agronomy 10
[11] Izzuddin M A, Hamzah A, Nisfariza M N and Idris A S 2020 Analysis of multispectral imagery from unmanned aerial vehicle (UAV) using object-based image analysis for detection of ganoderma disease in oil palm J. Oil Palm Res. 32 497–508
[12] Sankaran S, Mishra A, Ehsani R and Davis C 2010 A review of advanced techniques for detecting plant diseases Comput. Electron. Agric. 72 1–13
[13] Li X, Lee W S, Li M, Ehsani R, Mishra A, Yang C and Mangan R 2012 Spectral difference analysis and airborne imaging classification for citrus greening infected trees Comput. Electron. Agric. 83 32–46
[14] Santoso H, Gunawan T, Jatmiko R H, Darmosarkoro W and Minasny B 2011 Mapping and identifying basal stem rot disease in oil palms in North Sumatra with QuickBird imagery Precis. Agric. 12 233–48
[15] Tan K P, Kanniah K D and Cracknell A P 2013 Use of UK-DMC 2 and ALOS PALSAR for studying the age of oil palm trees in southern peninsular Malaysia Int. J. Remote Sens. 34
[16] Trisasongko B H, Paull D J, Griffin A L, Jia X and Panuju D R 2019 On the relationship between the circumference of rubber trees and L-band waves Int. J. Remote Sens. 40 6395–417

[17] Panuju D R, Paull D J, Griffin A L and Trisasongko B H 2021 Mapping rice growth stages employing MODIS NDVI and ALOS AVNIR-2 Remote Sensing and GIScience : Challenges and Future Directions ed P Kumar, H Sajjad, B S Chaudhary, J S Rawat and M Rani (Cham: Springer International Publishing) pp 185–203

[18] Liu M, Wang M, Wang J and Li D 2013 Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegarLiu M, Wang M, Wang J, Li D. 2013. Comparison of random Sensors Actuators B Chem. 177 970–80

[19] Santoso H, Tani H and Wang X 2017 Random Forest classification model of basal stem rot disease caused by Ganoderma boninense in oil palm plantations Int. J. Remote Sens. 38 4683–99

[20] Nasution T D S, Supriadi and Damanik M M B 2016 Surveying and mapping K nutrient status and C-organic on Oil Palm Plantations were attacked by Ganoderma in PT. PD. PATI Aceh Tamiang Agroekoteknologi 4 2238–44

[21] Tucker C J 1979 Red and photographic infrared linear combinations for monitoring vegetation Remote Sens. Environ. 8 127–50

[22] Gitelson A A, Kaufman Y J and Merzlyak M N 1996 Use of a green channel in remote sensing of global vegetation from EOS- MODIS Remote Sens. Environ. 58 289–98

[23] Wang F, Huang J and Chen L 2010 Development of a vegetation index for estimation of leaf area index based on simulation modeling J. Plant Nutr. 33 328–38

[24] Kaufman Y J and Tanre D 1992 Atmospherically resistant vegetation index (ARVI) for EOS-MODIS IEEE Trans. Geosci. Remote Sens. 30 261–70

[25] Huete A R 1988 A soil-adjusted vegetation index (SAVI) Remote Sens. Environ. 25 295–309

[26] Jordan C 1969 Derivation of leaf-area index from quality of light on the forest floor Ecology 50 663–6

[27] Liaghat S, Ehsani R, Mansor S, Shafri H Z M, Meon S, Sankaran S and Azam S H M N 2014 Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms Int. J. Remote Sens. 35 3427–39

[28] Breiman L 2001 Random Forests Mach. Learn. 45 5–32

[29] Santoso H, Tani H, Wang X, Prasetyo A E and Sonobe R 2019 Classifying the severity of basal stem rot disease in oil palm plantations using WorldView-3 imagery and machine learning algorithms Int. J. Remote Sens. 40 7624–46

[30] Sari M, Sonmez N K and Karaca M 2005 Relationship between chlorophyll content and canopy reflectance in Washington navel orange trees (Citrus sinensis (L.) Osbeck) Pak. J. Bot 37 1093–102

[31] Khalilia M, Chakraborty S and Popescu M 2011 Predicting disease risks from highly imbalanced data using random forest BMC Med. Inform. Decis. Mak. 11

[32] Trisasongko B H, Panuju D R, Paull D J, Jia X and Griffin A L 2017 Comparing six pixel-wise classifiers for tropical rural land cover mapping using four forms of fully polarimetric sar data Int. J. Remote Sens. 38 3274–93