RESEARCH PAPER

Analysis and prediction of rubber tree phenological changes during *Pestalotiopsis* infection using Sentinel-2 imagery and random forest

Yeni Herdiyeni [iD],[a,*] Muhammad Faishal Mumtaz,[a] Gibtha Fitri Laxmi,[a] Yudi Setiawan [iD],[b] Lilik Budi Prasetyo,[b] and Tri Rapani Febbiyanti[c]

[a]IPB University, Department of Computer Science, Faculty of Mathematics and Natural Sciences, Bogor, Indonesia

[b]IPB University, Department of Forest Resources Conservation and Ecotourism, Faculty of Forestry and Environment, Bogor, Indonesia

[c]Indonesian Rubber Research Institute Sembawa, Banyuasin, Indonesia

ABSTRACT. The fungus *Pestalotiopsis* sp. causes aberrant leaf fall, resulting in an annual reduction in rubber production. The loss of latex yield has a significant impact on the income of smallholder farmers. The objective of this study is to estimate NDVI using reflectance from the upper canopy based on Sentinel-2 NDVI time series data and daily climate data, analyze phenological changes of rubber trees using a harmonic model and time-lag cross-correlation, and develop a prediction model of vegetation indices using the random forest regressor (RFR) algorithm. The study was conducted on the rubber plantation of the Indonesian Rubber Research Institute in Sembawa, Indonesia. Three rubber clones aged 10 to 11 years were used to study the wintering trend. The study found that the overall trend of the NDVI decreased significantly from 2019 to 2022. The cycle of defoliation and refoliation changed due to disease and climate change. The time lag effects of vegetation index and climate variables are crucial for predicting vegetation dynamics. Increasing temperature and changing precipitation play significant roles in influencing pathogen incidence. The mean absolute percentage error (MAPE) was used to evaluate the trend of the vegetation indices. The RFR algorithm predicts the vegetation indices with a MAPE of 6.07%, 5.96%, and 8.18% of BPM24, GT1, and RRIC 100, respectively. Our finding indicated that the analysis and prediction model can be used to understand the phenological pattern and predict the wintering pattern to prevent disease spread and minimize latex yield loss due to infestation outbreaks.

1 Introduction

The *Pestalotiopsis* sp. fungal disease outbreak has significantly reduced the output of Indonesia's rubber plantations over the last 6 years (2016 to 2021).[1,2] *Pestalotiopsis* sp. has infected mature rubber trees on at least 0.4 million hectares. The spread of this disease afflicted South Kalimantan, Central Sulawesi, Lampung, Central Java, and East Java. In *Pestalotiopsis*, up to 75% to 90% of the defoliation is abnormal, and more than 25% of the latex yield is lost.[1]

*Address all correspondence to Yeni Herdiyeni, yeni.herdiyeni@apps.ipb.ac.id

Rubber cultivation in Indonesia is dominated by smallholder plantations. As a result, it has a negative impact on the well-being of rubber smallholders.

Pestalotiopsis is distinguished by brown spots on juvenile leaves that develop into dark brown spots and cause the leaves to fall prematurely. The newly developed leaves are smaller than ordinary leaves. The transitions from leaf defoliation to new leaf production (refoliation) are the primary phenological stages in rubber plantations. Plant phenology is the study of how vegetation develops throughout the year. Ground observation and remote sensing are the two methodologies used to study vegetation phenology.[3–5] On the ground, species-specific phenology reactions are found at limited scales. This procedure is time-consuming and expensive. Changes in vegetation can also be visible in high spatial and spectral resolution remote sensing imagery. Remote sensing has the potential to detect changes in vegetation features at numerous sizes, from the local to the continental and even the global.[6] Remote sensing has been used to collect data on plant biophysics, mostly on pathogen-induced physiological changes and diseases.[7–10] Previous study has shown that tracking the phenological stages of a rubber plantation necessitates observing the phenological dynamics of the rubber tree.[11–15] Based on phenology, it is feasible to investigate how a rubber plantation changes during the dry season, from defoliation to the growth of new leaves.[15]

The content of chlorophyll in leaves has been used to track changes in vegetation. The normalized difference vegetation index (NDVI) has been widely utilized and implemented to evaluate the chlorophyll content for vegetation development and health in areas with complete leaf coverage.[16] In rubber plantations, NDVI was found to be more susceptible to phenological fluctuations in canopy moisture and vegetation cover. According to the study, NDVI has been successfully employed to detect wintering in rubber plantations using MODIS images.[16]

Accurate time-series of vegetation indices are crucial for monitoring vegetation phenology in order to detect early stages of diseases and reduce production loss due to infestation outbreaks. Time series of NDVI from satellite data can be used to approximate phenological states and quantify overall vegetation behavior.[17] The time-series NDVI obtained from remote sensing pictures is commonly used in the extraction of vegetation phenology.[18–22] The prediction of vegetation dynamics using time series data is critical for disease control. Recent research has shown that random forest (RF) may accurately predict vegetation phenology.[23]

The study of the phenology of the rubber tree is critical for disease prevention. Zhai et al. studied whether prolonged defoliation enhanced powdery mildew disease infection rates and resulted in lower latex yield.[24] Gasparotto et al. discovered that the yearly defoliation and refoliation cycles that occur in rubber plants after 5 years of age have a considerable connection with foliar disease attacks.[25] Although phenological studies have risen in recent decades, we simply do not have any information on phenological studies of different rubber clones during *Pestalotiopsis* disease, particularly in tropical rubber plantations, based on Sentinel-2 imagery.

The main purpose of this study is to contribute to rubber disease monitoring by observing the phenological changes of rubber plantations, specifically the pattern of the wintering period based on Sentinel-2 NDVI time series data and local weather station data from 2019 to 2022 during *Pestalotiopsis* infection. The goals of this study are as follows: (1) estimating NDVI using reflectance from the upper canopy based on Sentinel-2 imagery across three rubber clones (BPM 24, GT 1, and RRIC 100); (2) analyzing phenological changes of rubber trees on the wintering period during *Pestalotiopsis* infection using a harmonic model and time-lag cross-correlation; and (3) developing a prediction model of vegetation indices using the random forest regressor (RFR) algorithm.

2 Material and Methods

2.1 Experimental Locations and Workflow

The study area of this research is the rubber plantation of Indonesian Rubber Research Institutes (IRRI) in Sembawa, South Sumatra, Indonesia. The fields have been infected by *Pestalotiopsis* diseases since 2016 with varying degrees of canopy infection severity. Three fields with different rubber tree clones, namely BPM 24, RRIC 100, and GT 1, and aged 10 to 11 years were selected

as experimental areas in this research. Figure 1 and Table 1 present the location and digital elevation model image of the study area.

The workflow of the proposed work is shown in Fig. 2. The method starts with data acquisition, pre-processing, NDVI calculation, imputation and smoothing, analysis of phenological changes, prediction of vegetation indices, and evaluation.
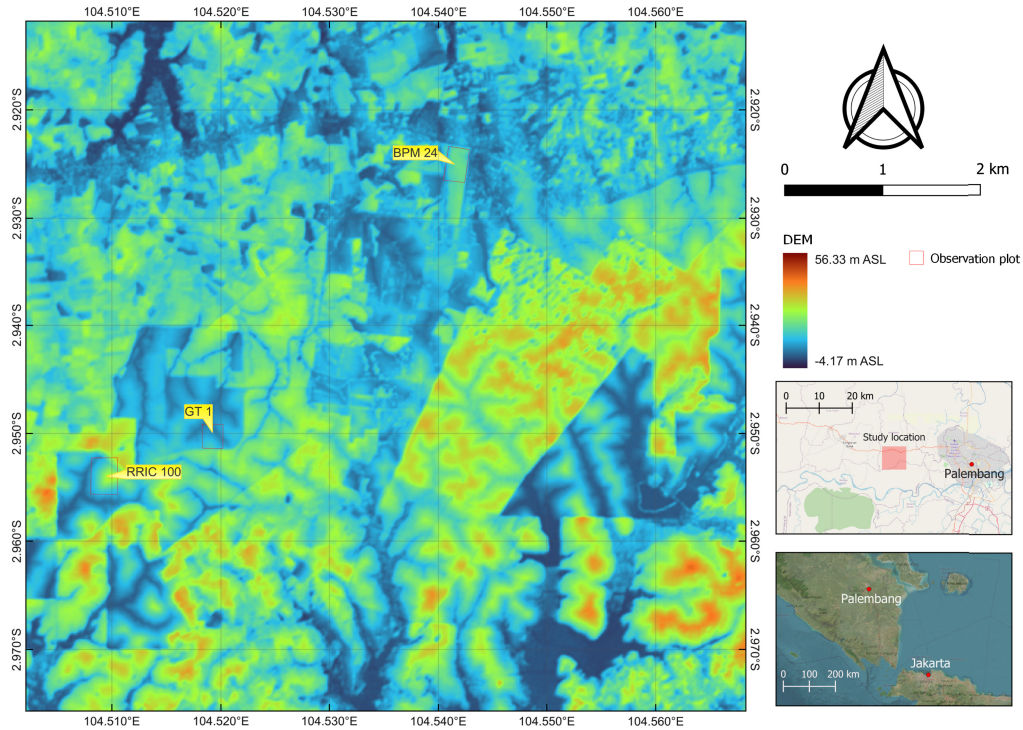

Fig. 1 Location of the study area.

Table 1 Extent of each field for three rubber tree clones.

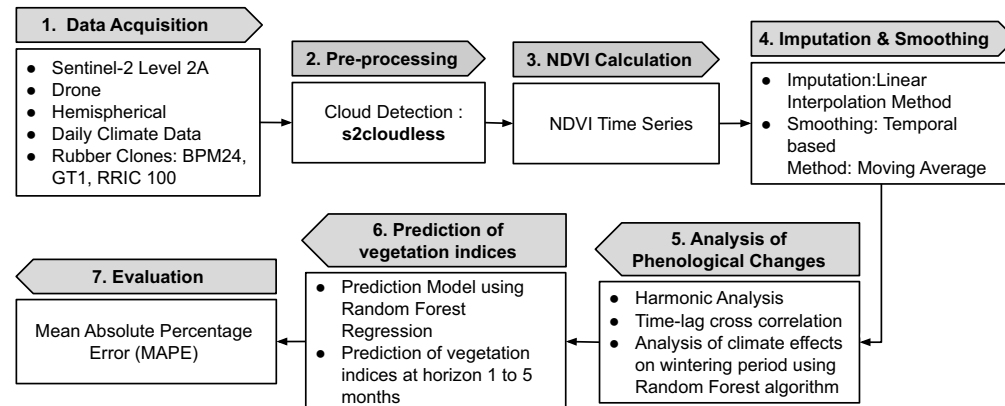| Clone name	Left	Top	Right	Bottom
BPM 24	104°32′27.635″ E	2°55′27.591″ S	104°32′30.594″ E	2°55′32.0406″ S
RRIC 100	104°30′31.456″ E	2°57′11.991″ S	104°30′36.474″ E	2°57′16.182″ S
GT 1	104°31′6.845″ E	2°56′58.121″ S	104°31′12.169″ E	2°57′3.355″ S

Fig. 2 General workflow of the proposed method: (1) data acquisition, (2) pre-processing, (3) NDVI calculation, (4) imputation and smoothing, (5) analysis of phenological changes, (6) prediction of vegetation indices, and (7) evaluation.

2.2 Data Acquisition

In this study, we obtained a 4-year time series data from the Sentinel-2 level 2A product of the study areas from January 1, 2019, to December 31, 2022. The 291 imagery data were acquired from the Google Earth Engine (GEE) platform. Sentinel-2 provides reflectance data for 13 bands and has a 5-day temporal resolution. For this study, we used four spectral bands from a top-of-atmosphere product with a spatial resolution of 10 m: B02 (blue), B03 (green), B04 (red), and B08 (infrared). We also collected UAV multispectral data and digital hemispherical data during disease outbreak as the ground truth data for disease severity at two time periods, between January 18–19, and June 15–16, 2022. The use of UAV data enables a broader understanding of the affected area, providing increased resolution and enhanced accuracy through hemispherical data. These data were used by experts to assess the health of vegetation and determine the severity of *Pestalotiopsis* disease. Daily climate data were obtained from local weather stations in the rubber plantation and the Indonesian Central Bureau of Statistics and the Indonesian Meteorology, Climatology, and Geophysical Agency from 2019 to 2022.[26]

2.3 Pre-Processing Sentinel-2 Imagery

Cloud cover is a common problem in evaluating time-series data using optical imaging. Cloud removal is vital for the subsequent data analysis and to increase the integrity and availability of the remote sensing data. There are three approaches of cloud removal, i.e., spatial information-based method, temporal information-based methods, and multi-source auxiliary information-based method.[27] Many studies on cloud removal have been conducted.[28–31]

   To eliminate cloud cover in this experiment, we used *s2cloudless*, a machine learning-based cloud identification technology developed by Sinergise's EO Research team. This technique leverages the light GBM algorithm, which was trained on a massive dataset of cloudy and non-clouded samples from around the world.[32] Before applying the gradient boost-based technique, all bands are upsampled to 10 m resolution using bilinear interpolation. The algorithm employs Sentinel-2 Level-1C TOA reflectance values from the following 10 bands as input: B01, B02, B04, B05, B08, B8A, B09, B10, B11, B12; and the output is a cloud probability map. In this study, we used a threshold value of 0.4 to reduce cloud omission errors.

2.4 Calculation of NDVI

To explore the rubber trees phenology stages and define vegetation behavior during *Pestalotiopsis* infestation, we studied the NDVI-time series data acquired from Sentinel-2 datasets. The time series graph was created using the NDVI from random sites in the three rubber fields. The NDVI is determined using the equation as follows:

$$NDVI = (NIR - RED)/(NIR + RED), \tag{1}$$

where NIR is the reflectance of NIR radiation and RED is the reflectance of visible RED energy. The index for Sentinel-2 can be derived as the ratio of the difference and total of reflectance in NIR (B8) and red (B4) areas, as shown as:[33]

$$NDVI = (B8 - B4)/(B8 - B4). \tag{2}$$

2.5 Imputation and Smoothing NDVI Time Series

The missing information generated by cloud causes spatial and temporal gaps in satellite earth observation data, as well as biases in subsequent image processing and application. Various methods for dealing with missing data have been proposed. Methods are classified into four types: spatial-based methods, spectral-based methods, temporal-based methods, and hybrid methods.[34] In this study, we used a temporal-based method or "gap filling" technique. Recent temporal techniques, such as data fusion,[35,36] pixel unmixing,[37] data interpolation,[38] or best-pixel selection,[39,40] have been examined for large-area land cover mapping. To obtain the gap-filled satellite image, we used the linear interpolation method in this work. The linear interpolation function equation is as follows:[41]

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0), \tag{3}$$

where x is the independent variable, x_1 and x_0 are known values of the independent variable, and $f(x)$ is the value of the dependent variable for a value x of the independent variable.

The noise of the signals received by the satellite sensors can occur due to cloud cover, atmospheric disturbance, sensor failure, and so on.[42] The time series smoothing method can be used to reduce noise and assess plant phenological trends.[42] Smoothing methods were investigated by Atkinson et al. in reconstructing consecutive time series of vegetation indices to estimate vegetation phenology.[43] There were three types of smoothing methods:[44] temporal methods, frequency methods, and hybrid methods. In this study, we used temporal-based method, a moving average (MA) filter for smoothing NDVI time series data. The MA filter reconstructs NDVI time series by replacing the central point with the average of all points in an equation-based moving window of fixed size. The advantage of this strategy is that the seasonal peak of the curve is preserved.

2.6 Analysis of Phenological Changes

The analysis of phenological changes of NDVI was investigated using harmonic analysis. The harmonic analysis of NDVI time series aims to characterize the spectral behavior of NDVI by decomposing the time series into a sequence of basic sinusoids and a set of sine or cosine terms with unique coefficients and intercept (harmonic terms).[45] Time-lag cross-correlation analysis was used to understand the relationship and potential lag (lead effects) between NDVI and climate variables. Cross-correlation measures the similarity between two time series as a function of the time lag applied to one of them. Time lag indicates the time delay between precipitation and temperature changes and their impact on NDVI.

2.7 Prediction of Vegetation Indices

In this study, we analyzed the relationship between NDVI variations and climatic variables to understand the impact of the *Pestalotiopsis* diseases on vegetation dynamics in a rubber plantation using RFR. RFR was first proposed by Breiman (1996) as an ensemble learning method that can be used in both classification and regression task.[46] RFR is a non-linear ensemble technique that focuses solely on decision trees for classification or regression. The concept behind RF for time series is to replace the standard bootstrap with a dependent bootstrap in order to subsample time series during the tree construction phase in order to account for time dependence.[46]

2.8 Evaluation

To validate the performance of the model, multi-step-ahead monthly NDVI forecasts were simulated to walk-forward validation. The walk-forward validation was utilized to improve the model's precision. The dataset time series was divided into training, validation, and testing sets, with 90% of the dataset used for training and validation, and 10% retrained for testing. Means absolute percentage error (MAPE) performance evaluation was conducted to validate the accuracy of the proposed model. The following expression is the MAPE equation:

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{E_t - A_t}{A_t} \right|, \tag{4}$$

where E_t is the predicted value for t'th, and the A_t is the element is the actual value for t'th.

3 Results and Discussion

Understanding and predicting the leaf phenology changes of rubber trees can help control diseases. The pattern of defoliation and refoliation of rubber trees (wintering period) is used to understand disease spread and severity, as well as to plan or decide on chemical spraying to control diseases.[45] Wintering happens when rubber developed an adaptation to reduce transpiration due to water scarcity during the dry season. Wintering patterns are well known to have an impact on yield. In this study, we analyzed phenological changes during through wintering period during *Pestalotiopsis* infection using Sentinel-2 NDVI time series data and daily climate data from 2019 to 2022.

3.1 Environmental Factors Affecting *Pestalotiopsis* Disease

The incidence and severity of *Pestalotiopsis* sp. disease are intricately correlated with various environmental factors that influence the fungus's growth, spread, and impact on plant populations.[47] *Pestalotiopsis* sp. prefers specific temperatures for growth and reproduction. Extreme temperatures cause stress in plants, making them more susceptible to infections.[48] Temperature fluctuations can also affect the fungus's growth rate and development, as well as the host plants'. High humidity can create an ideal microclimate for fungal growth. Humidity, along with temperature, affects spore production and dispersal. *Pestalotiopsis* sp. thrives in humid environments, as do many other fungi.

Environmental stressors, such as intense rainfall or flooding, have the potential to accelerate disease spread by creating favorable conditions for spore dispersal and germination.[48,49] In contrast, prolonged periods of drought have the potential to diminish the vitality of plants and render them more vulnerable to infections.[50,51]

The vegetation index is an effective tool for assessing plant health in relation to *Pestalotiopsis* disease. A declining vegetation index may indicate stress on plant populations, which could be caused by *Pestalotiopsis* spp. infection. Monitoring this index enables the early detection of potential outbreaks. A thorough understanding of the interactions between temperature, precipitation, vegetation index, and other environmental factors is required for predicting, preventing, and managing *Pestalotiopsis* disease in rubber plantations. The analysis of the relationships between these variables allows for the identification of patterns and correlations, which aids in the prediction of *Pestalotiopsis* sp. outbreaks, the implementation of preventive measures, and the development of disease management strategies in affected ecosystems.[47]

3.2 Phenological Changes Based on Harmonic Analysis of NDVI Time

In this study, we observed three rubber clones (GT 1, BPM24, and RRIC 100) aged 10 to 11 years to study the NDVI trend. Figures 3(a)–3(c) show a harmonic model of NDVI time series from 2019 to 2022 for three rubber clones. We found that the overall trend of the NDVI decreased significantly from 2019 to 2022 for all clones.

Pestalotiopsis was reported in rubber plantations in North Sumatra and South Sumatra since 2016.[1] Figure 4 depicts defoliation and refoliation trends of each clone. According to previous research phenology stage of defoliation and refoliation period for rubber wintering occurs in the dry season, which in Sumatra is between July and September[16] (as shown in Fig. 4 in gray area). Defoliation takes place over a 4-week span from July to August, signifying the conclusion of the growing season. Simultaneously, it marks the onset of the subsequent growing season, transitioning into the refoliation period spanning August to September. During 2019 to 2022, we found that defoliation and refoliation events shifted due to *Pestalotiopsis* disease. Early wintering was reported in all clones. The first defoliation occurred in March to July 2019, followed by secondary leaf fall in November and January 2020. In 2020 and 2021, the defoliation occurred repeatedly or extensively in all clones, indicating physiological stress, nutrient depletion, and a compromised defense system. These factors create favorable conditions for pests and pathogens to thrive, leading to increased vulnerability and the potential for further damage to the trees.

In 2022, the NDVI gradually decreased and defoliation significantly and occurred earlier in March. All clones suffered from extensive defoliation between March and July in 2022. According to weather statistics data in Fig. 5, in 2022, precipitation increased and caused a decrease in minimum temperature compared to 2019 to 2021. The increase in precipitation can create a cascade of unfavorable conditions for rubber trees, including the spread of fungal diseases, waterlogged soil, nutrient deficiencies, disruption of photosynthesis, pest infestations, and physiological stress. These factors collectively contribute to the extensive defoliation observed in rubber trees during periods of heavy rainfall and this can be reflected in temporal NDVI profiles with shifts in the timing of NDVI peaks and valleys.

The phenology of rubber trees is affected by the period of defoliation. A longer defoliation period would consume more conserved carbohydrates, reducing available reserves for disease defense and latex production.[24] This condition is associated with significant defoliation and decrease in rubber yield. The shifts in the starting and ending periods of the annual dry season and wintering are likely the result of a complex interplay between *Pestalotiopsis* disease and broader climate change impacts.
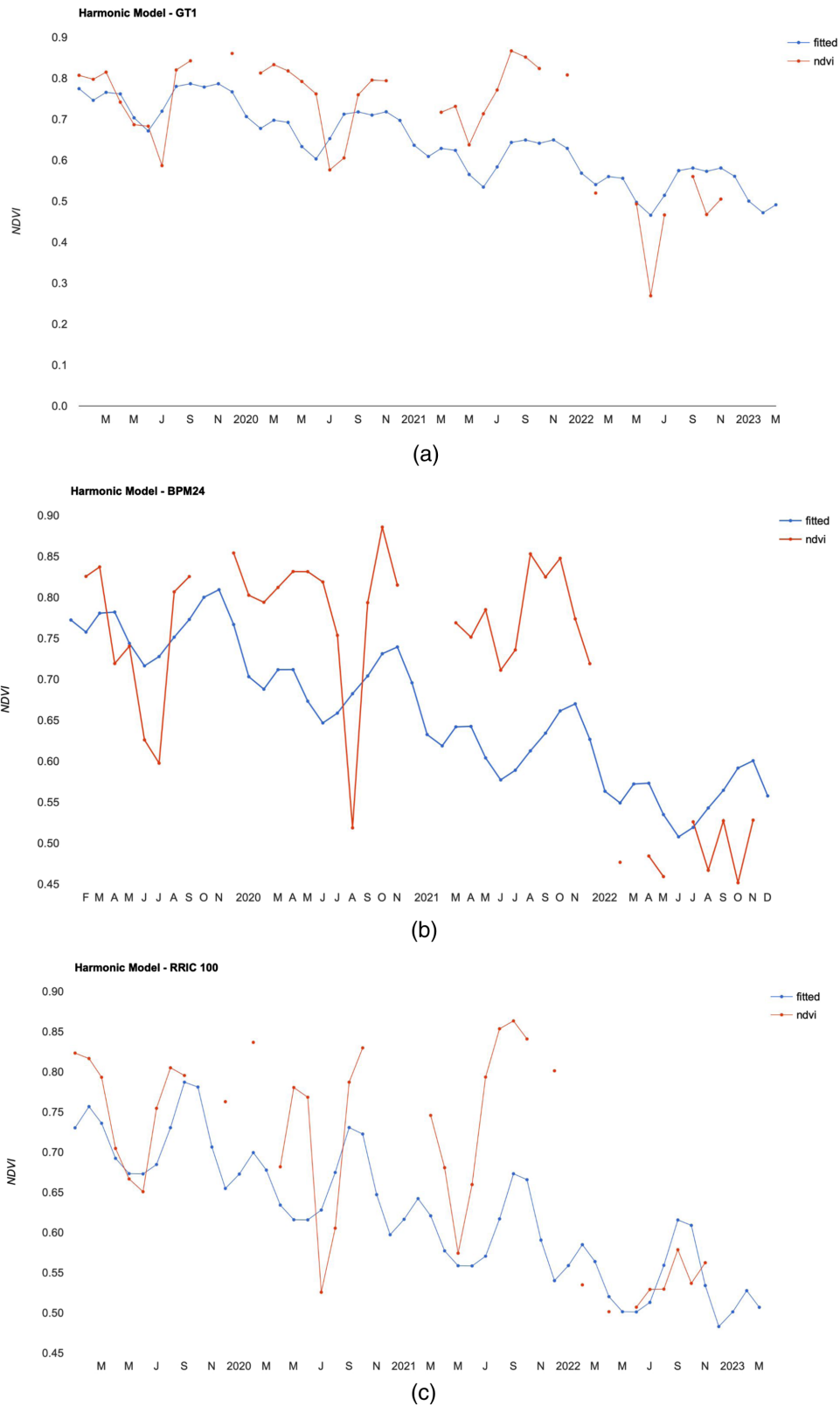
Fig. 3 Harmonic model of NDVI time series over four years (from 2019 to 2022) for three clones, (a) GT1, (b) BPM24, and (c) RRIC 100. The red curve illustrates the fitted model, while the blue points denote the actual NDVI data.
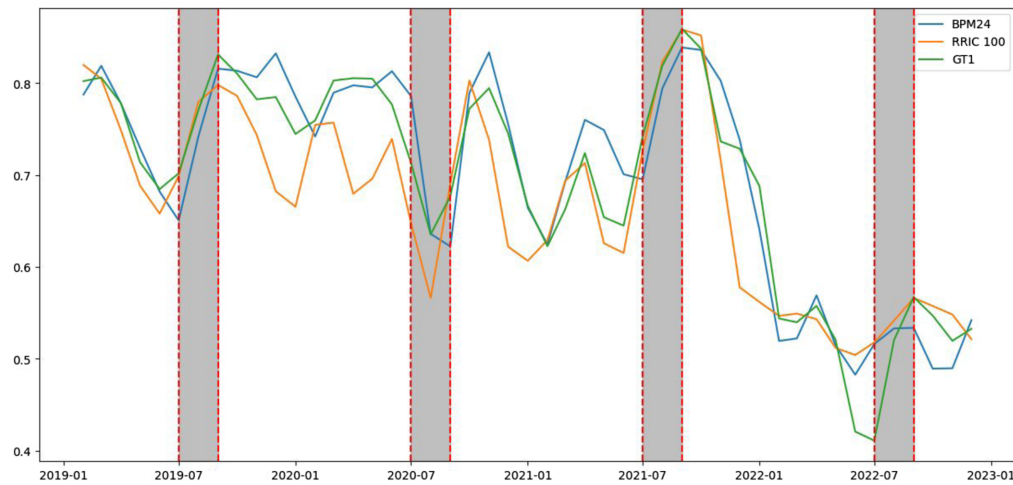
Fig. 4 Defoliation and refoliation trends of several rubber clones (BPM 24, GT 1, and RRIC 100) with *Pestalotiopsis* infection in 2019 to 2022. The gray area denotes the annual wintering period in Sumatra (between July and September).
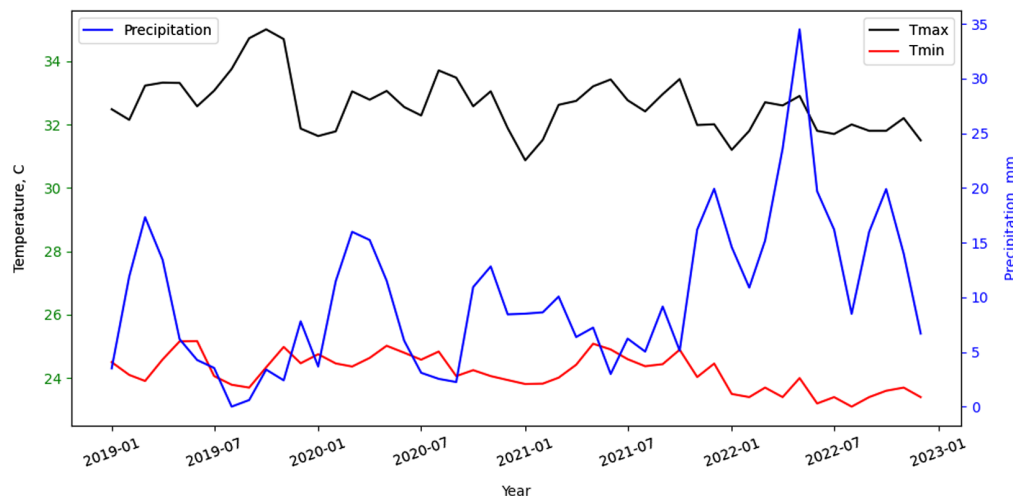

Fig. 5 Temperature (Tmin and Tmax) and precipitation trends during *Pestalotiopsis* outbreak from 2019 to 2022.

The proliferation of *Pestalotiopsis* leaf disease has caused extensive defoliation. This symptom indicates the presence of *Pestalotiopsis* disease in all rubber clones. To prove the occurrence of disease in rubber plantations, in 2022 we collected UAV multispectral data and digital hemispherical data as the ground truth data for disease severity at two time periods, between January 18–19, and June 15–16, 2022. Pathologist experts from IRRI also conducted research to determine the severity of *Pestalotiopsis* disease in rubber plantations. Figure 6 shows the vegetation index maps of disease severity at different fields (BPM 24, RRIC 100, and GT 1) on January 18, 2022. The colors green, yellow, and red on the map represent the values of NDVI which are high, moderate, and low, respectively. Higher values indicate healthy plants and lower values indicate a plant infested with pests, damaged plants, or unhealthy plants.

3.3 Analysis and Prediction of NDVI Based on Climate Variable Time Series Using Random Forest Algorithm

The values of NDVI, temperatures, humidity, precipitation, radiation, and wind speed were used to assess phenological changes in response to disease outbreaks and climate changes for three rubber clones. We applied the augmented Dickey–Fuller test (ADF test) to test the stationary time series. If the p-value is >0.05, then the data series is non-stationary. Table 2 shows that the
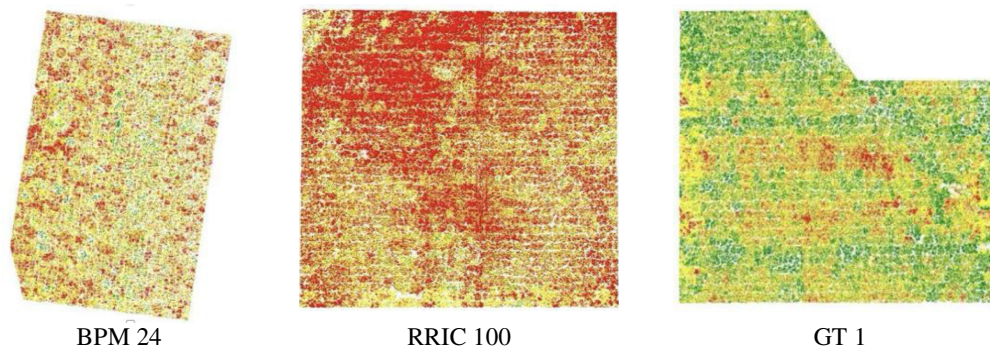
BPM 24   RRIC 100   GT 1

Fig. 6 Time-lag cross correlation between NDVI and climate variables (precipitation and temperature).

p-values of NDVI for all clones (GT1, BPM24, RRIC), precipitation, maximum temperature, humidity, and radiation are greater than the significance level of 0.05, and the ADF statistic is higher than any of the critical values. Therefore, the sequence of NDVI and climate data were non-stationary.

We applied time-lag cross-correlation analysis to understand the relationship and potential lag (lead effects) between NDVI and climate variables. We found a time lag which indicates the time delay between precipitation and temperature changes and their impact on NDVI. Figure 7 shows the lag value associated with the peak of the correlation function of precipitation, temperature, and NDVI. Results show that vegetation has a 1- to 3-month lag response to precipitation and temperature anomalies. It means the precipitation and temperatures highly correlate with a time shift of 1 to 3 months. A negative peak cross-correlation between NDVI and precipitation and a positive peak cross-correlation between NDVI and temperature anomaly correlation are crucial for assessing the impact of changing climate patterns. Anomaly correlation between NDVI and climate variables indicates how variations in climate patterns affect vegetation growth and disease prevalence. The time lag between climatic conditions and changes in NDVI reflects the health conditions of vegetation caused by *Pestalotiopsis* disease. The significant reduction in NDVI values resulting from disease outbreaks and climatic conditions provides evidence for the development of *Pestalotiopsis*.

We applied the lagging time series to the features for predicting the NDVI changes using RFR based on time-lag cross-correlation of temperature and precipitation. Table 3 shows the features of the prediction model.

In order to evaluate the performance of the predictive model, we utilized time series cross-validation, a process that involves conducting multiple validation iterations. The prediction model was built using data from 2019 to 2023. The entire time period was divided into training and testing sets. The expanding window walk-forward validation method, specifically a fivefold time series cross-validation, was used in our evaluation, as shown in Fig. 8. As the cross-validation process progressed, the training data expanded to encompass all data that had preceded it, while the size of the testing data remained constant. The objective of walk-forward validation was to enhance the accuracy of the model in predicting the initiation of defoliation, with forecasting intervals varying between 1 and 5 months.

Before starting the model's training and testing phases with the RFR model, we used randomized search and grid search hyperparameter tuning. Randomized search is used as the first step in hyperparameter optimization. It aids in quickly gaining an understanding of the hyperparameter space and identifying promising areas for further investigation through grid search. The training set was subjected to grid-search time-series cross-validation in order to identify optimal hyperparameters from the search space and to ensure unbiased tuning. We ran a randomized search for each fold in the time series cross-validation. This experimental setup was used for three different rubber clones. Table 4 shows the description and optimal hyperparameters used in the RFR model for each clone.

Figures 9(a)–9(c) show the predicted NDVI results for each clone, with time horizons ranging from 1 to 5 months. The performance of the prediction model is measured by the mean absolute percentage error (MAPE) which yields the following results for each rubber clone:

Table 2 ADF test to test the stationary of time series.

|	GT1	BPM24	RRIC	Precipitation	Tmin	Tmax	Tavg	Diff_temp	Humidity	Radiation	Wind speed
Test statistic	-1.166	-0.923	-1.134	-0.875	-2.905	-1.142	-1.602	-3.169	-0.669	-1.408	-2.944
P-value	0.688	0.780	0.701	0.796	0.045	0.698	0.483	0.022	0.855	0.579	0.040
Lags used	2.000	2.000	5.000	8.000	1.000	10.000	8.000	8.000	8.000	9.000	7.000
No. of observations	44.000	44.000	41.000	38.000	45.000	36.000	38.000	38.000	38.000	37.000	39.000
Critical value (1%)	-3.589	-3.589	-3.601	-3.616	-3.585	-3.627	-3.616	-3.616	-3.616	-3.621	-3.610
Critical value (5%)	-2.930	-2.930	-2.935	-2.941	-2.928	-2.946	-2.941	-2.941	-2.941	-2.944	-2.939
Critical value (10%)	-2.603	-2.603	-2.606	-2.609	-2.602	-2.612	-2.609	-2.609	-2.609	-2.610	-2.608
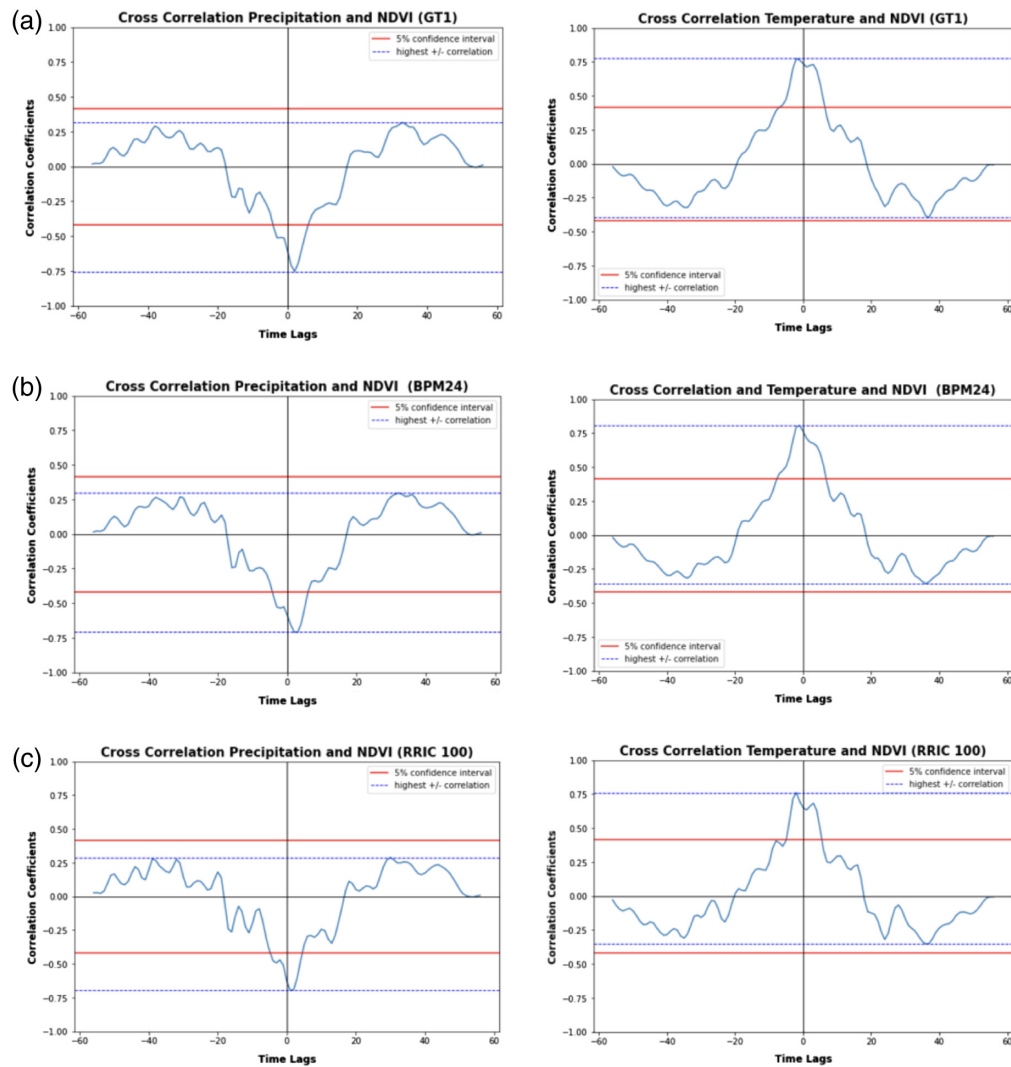
Fig. 7 Time-lag cross correlation between NDVI and climate variables (precipitation and temperature) for (a) GT, (b) BPM24, and (c) RRIC 100.

Table 3 List of features for prediction model.

Features	Description	Features	Description
Precipitation	Precipitation	lag_pre2	Precipitation value 2 months before (lag features)
Radiation	Radiation	lag_rad2	Radiation value 2 months before (lag features)
Humidity	Humidity	lag_hum2	Humidity value 2 months before (lag features)
Wind Speed	Wind speed	lag_diff_temp2	diff_temp value 2 months before (lag features)
diff_temp	Difference maximum temperature and minimum temperature	lag_win2	Wind speed value 2 months before (lag features)
pre_temp	Precipitation/temperature	lag_pre_temp2	pre_temp value 2 months before (lag features)
lag_ndvi1	NDVI value 1 month before (lag features)	month	Month of NDVI

Table 3 (*Continued*).

Features	Description	Features	Description
lag_pre1	Precipitation value 1 month before (lag features)	rolling_mean_ndvi	mean of ndvi value from lag features
lag_rad1	Radiation value 1 month before (lag features)	rolling_mean_diff_temp	mean of diff_temp value from lag features
lag_hum1	Humidity value 1 month before	rolling_mean_pre	Mean of precipitation value from lag features
lag_diff_temp1	diff_temp value 1 month before	rolling_mean_pre_temp	Rolling mean lag features of pre_temp
lag_win1	Wind speed value one month before	rolling_mean_rad	Rolling mean lag features of radiation
lag_pre_temp1	pre_temp 1 month before (lag features)	rolling_mean_hum	Rolling mean lag features of humidity
lag_ndvi2	NDVI value 2 month before (lag features)	rolling_mean_win	Rolling mean lag features of wind speed
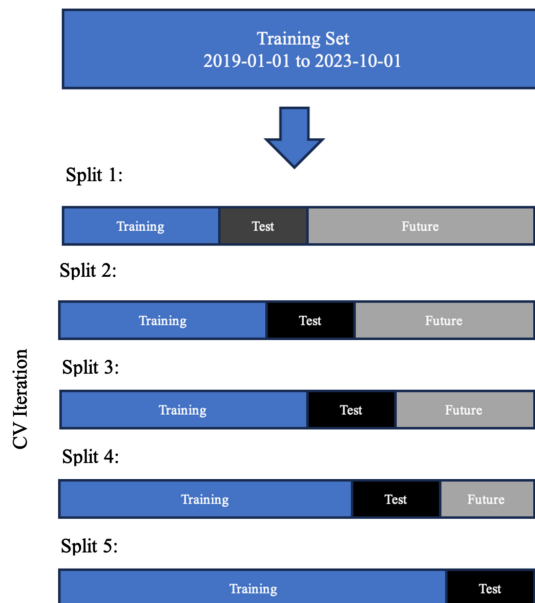
Fig. 8 Illustration of time series cross-validation.

Table 4 The description and optimal hyperparameter used in the RF regression model.

| Hyperparameter	Description	RRIC 100	BPM24	GT1
max_depth	The maximum depth of the tree	10	40	5
max_features	Number of random features sampled per node	30	30	40
min_samples_leaf	The minimal number of data points required in order to create a leaf	5	2	5
min_samples_split	The minimal number of data points required to split an internal node	4	4	2
bootstrap	Whether to train in bootstrap samples or on the full train set	True	True	True
n_estimators	The number of trees in the forest	25	25	25

(a)

(b)

(c)
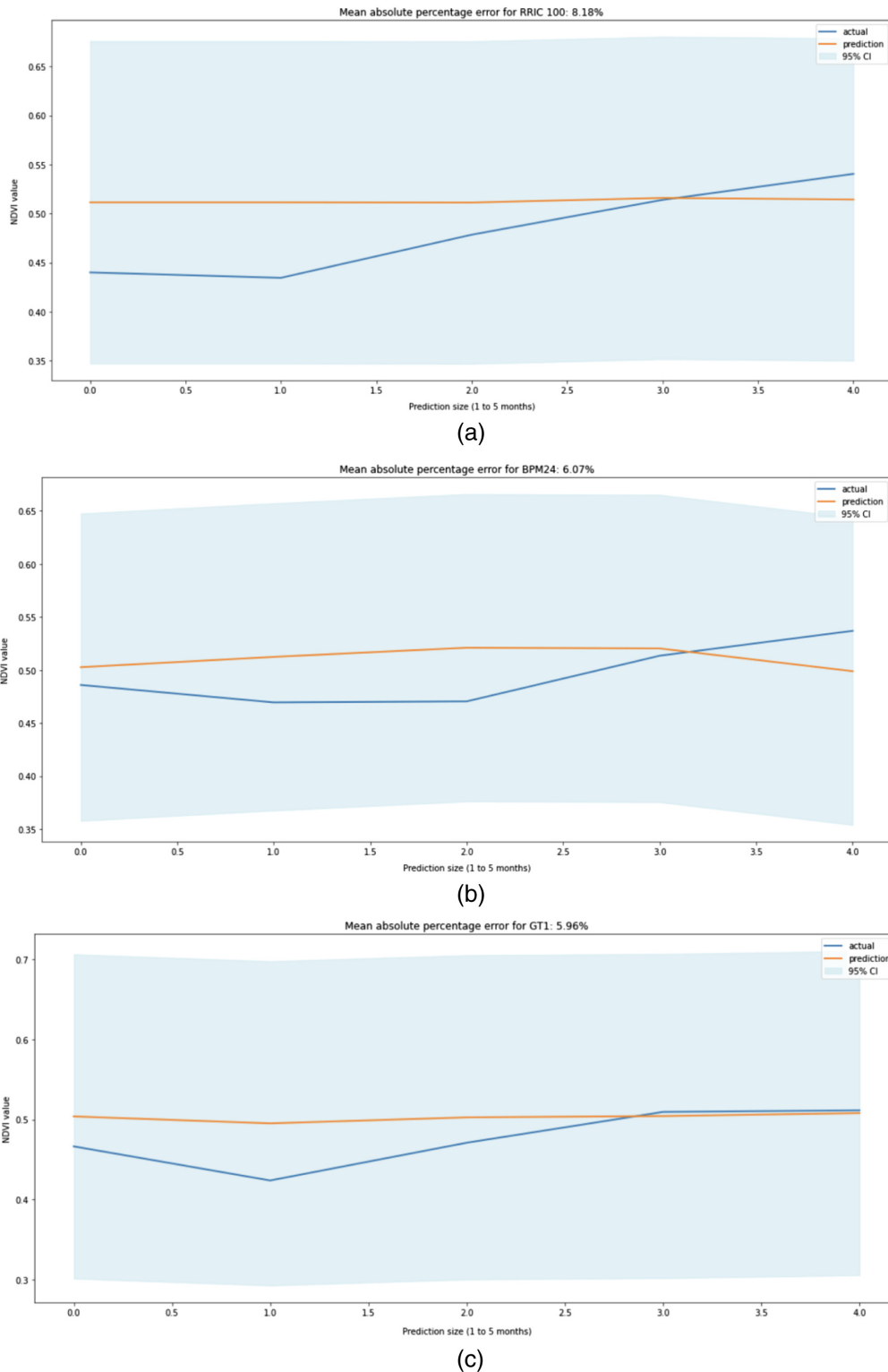
Fig. 9 The predicted result of NDVI with horizons of 1 to 5 months for (a) RRIC 100, (b) BPM24, and (c) GT1 clones.

8.18% for RRIC 100, 6.07% for BPM24, and 5.96% for GT1. The model had accuracies of 90.80%, 93.65%, and 93.35% for RRIC 100, BPM24, and GT1, respectively.

Figure 10 depicts the important variables in the correlation between NDVI and climate variables for individual rubber clones using RFR model. Incorporating time lag effects of vegetation and climate variables is especially useful in predictive analysis, particularly when vegetation
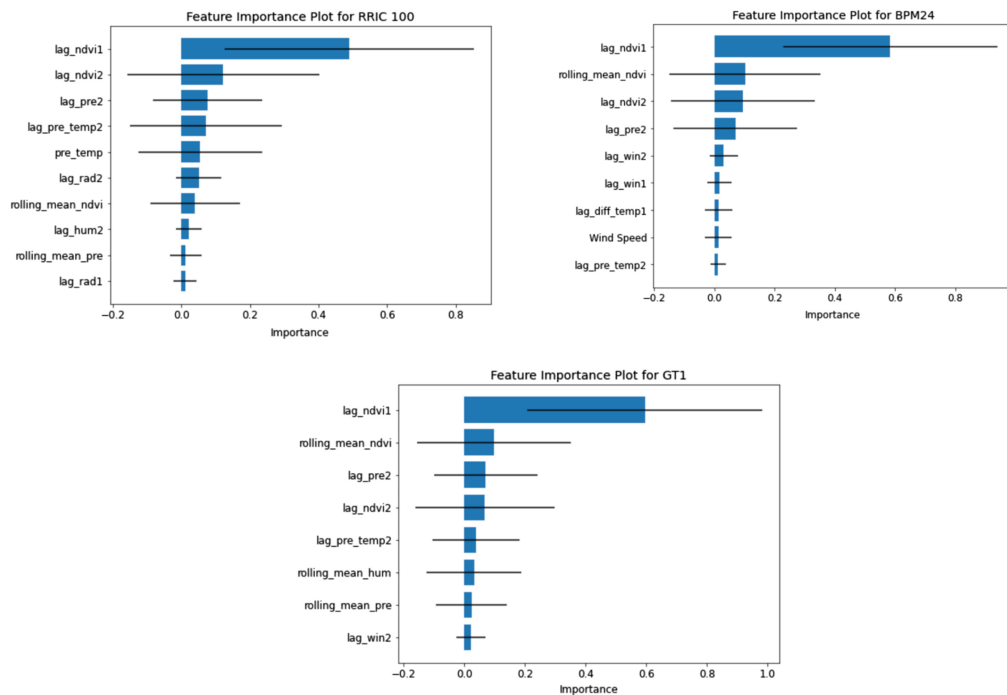
Fig. 10 Importance of variables in relationship between NDVI and climate variables.

responses to changing conditions follow temporal patterns. Our findings show that time lag effects recognize that the current NDVI value is influenced not only by the immediate past value but also by values from specific time intervals prior to the current observation. In this study, the time-lag effects of NDVI (*lag_ndvi1, lag_ndvi2, rolling_mean_ndvi*) and climate variables, such as precipitation (*lag_pre2, rolling_mean_pre*), temperature (*lag_diff temp1*), radiation (*lag_rad1, lag_rad2*), humidity (*lag_hum2*), and wind speed (*wind_speed, lag_win1, lag_win2*) influenced the prediction of NDVI values in time series data. Significantly, the 1-month interval preceding NDVI features (*lag_ndvi1*) emerges as the most important NDVI predictor for all rubber clones.

Precipitation and temperature have a much greater impact on the phenological pattern of all rubber clones than humidity, wind speed, and radiation. Important characteristics for NDVI value prediction include the significance of time-lag effects, particularly precipitation (*lag_pre2*), the variance in maximum and minimum temperatures (*diff_temp, lag_diff_temp2*), and the ratio of precipitation to temperature (*pre_temp, lag_pre_temp2*). The ratio of precipitation to temperature offers valuable insights into the equilibrium between temperature and water availability, a critical determinant in evaluating potential stress on vegetation. A disparity in temperature and precipitation may indicate water stress, which has negative impacts on plant development and overall ecosystem health. Furthermore, the combination of increased precipitation and lower temperatures creates an environment conducive to the growth, spread, and establishment of pathogenic microorganisms, particularly fungi. This phenomenon contributes to an increased prevalence of diseases among rubber trees. It is concluded that the temporal lag effects of climate variables and the vegetation index are significant in the prediction of forthcoming vegetation dynamics.

The results show the random-forest-based classification was highly effective in modeling the relationship between NDVI and climate variables. We found a strong correlation between NDVI and climate variables. This correlation indicates that these variables can be used to determine the vegetation cover and to assess phenological changes in response to disease outbreaks. The analysis indicated that the outbreak of *Pestalotiopsis* disease and the weather pattern have been confirmed to influence the pattern and timing of defoliation and refoliation of rubber trees (wintering period). We conclude that the analysis of the phenological pattern and the proposed prediction model can be used to forecast the early wintering period of rubber trees to prevent the spread of disease and minimize latex yield loss.

4 Conclusion

During the *Pestalotiopsis* disease outbreak, the phenological pattern of rubber trees changed. Disease and climate change have shifted the defoliation and refoliation cycles in rubber plantations. Structure of temporal sequences RFR was used in conjunction with NDVI data and climate variables to forecast the phenological stages or wintering pattern of rubber trees. Time lag effects in both climate variables and vegetation indices are critical for predicting future vegetation dynamics. The time lag between temperature and precipitation changes and their impact on the NDVI was determined to be 1 to 3 months. By utilizing this methodology, complex correlations between historical and current NDVI values can be captured, thereby improving the precision of the forecasts. MAPE was applied to evaluate of the NDVI changes with horizon of 1 to 5 months. RFR algorithm exhibits good performance in predicting the NDVI changes for three rubber clones, The MAPE of prediction model for each rubber clone are 8.18%, 6.07%, and 5.96% for RRIC 100, BPM24, and GT1, respectively. The investigation of the phenological pattern and the prediction of the early stages of rubber tree defoliation is essential for disease spread prevention and minimize latex yield loss.

References

1. N. Damiri et al., "*Pestalotiopsis sp.* infection causes leaf fall disease of new arrivals in several clones of rubber plants," *Biodiversitas* **23**(8), 3943–3949 (2022).
2. S. S. S. Aliya et al., "Colletotrichum siamense and *Pestalotiopsis jesterias* potential pathogens of new rubber leaf spot disease via detached leaf assay," *J. Rubber Res.* **25**, 195–212 (2022).
3. K. K. Liyanage et al., "Evaluation of key meteorological determinants of wintering and flowering patterns of five rubber clones in Xishuangbanna, Yunnan, China," *Int. J. Biometeorol.* **63**, 617–625 (2019).
4. M. Broich et al., "Land surface phenological response to decadal climate variability across Australia using satellite remote sensing," *Biogeosciences* **11**, 5181–5198 (2014).
5. S. Ren et al., "Diverse responses of vegetation phenology to climate change in different grasslands in Inner Mongolia during 2000-2016," *Remote Sens.* **10**(2), 17 (2017).
6. J. Zhang et al., "Detecting powdery mildew of winter wheat using leaf level hyperspectral measurements," *Comput. Electron. Agric.* **85**, 13–23 (2012).
7. J. Zhang et al., "Monitoring plant diseases and pests through remote sensing technology: a review," *Comput. Electron. Agric.* **165**, 104943 (2019).
8. A. K. Mahlein et al., "Development of spectral indices for detecting and identifying plant diseases," *Remote Sens. Environ.* **128**, 21–30 (2013).
9. J. Behmann et al., "A review of advanced machine learning methods for the detection of biotic stress in precision crop protection," *Precis. Agric.* **16**, 239–260 (2015).
10. J. C. Marín-Ortiz et al., "Linking physiological parameters with visible/near-infrared leaf reflectance in the incubation period of vascular wilt disease," *Saudi J. Biol. Sci.* **27**, 88–99 (2020).
11. J. Dong et al., "Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery," *Remote Sens. Environ.* **134**, 392–402 (2013).
12. H. Fan et al., "Phenology-based vegetation index differencing for mapping of rubber plantations using Landsat OLI data," *Remote Sens.* **7**(5), 6041–6058 (2015).
13. B. Chen et al., "Mapping tropical forests and deciduous rubber plantations in Hainan Island, China by integrating PALSAR 25-m and multi-temporal Landsat images," *Int. J. Appl. Earth Observ. Geoinf.* **50**, 117–130 (2016).

14. D. L. Zhai et al., "Comparison of pixel- and object-based approaches in phenology-based rubber plantation mapping in fragmented landscapes," *Remote Sens.* **10**(1), 44 (2018).

15. C. W. Xiao, P. Li, and Z. M. Feng, "Monitoring annual dynamics of mature rubber plantations in Xishuangbanna during 1987–2018 using Landsat time series data: a multiple normalization approach," *Int. J. Appl. Earth Observ. Geoinf.* **77**, 30–41 (2019).

16. F. A. Azizan et al., "Using multi-temporal satellite data to analyse phenological responses of rubber (*Hevea brasiliensis*) to climatic variations in South Sumatra, Indonesia," *Remote Sens.* **13**(15), 2932 (2021).

17. X. Huang et al., "The optimal threshold and vegetation index time series for retrieving crop phenology based on a modified dynamic threshold method," *Remote Sens.* **11**(23), 2725 (2019).

18. A. Buyantuyev and J. Wu, "Urbanization diversifies land surface phenology in arid environments: interactions among vegetation, climatic variation, and land use pattern in the Phoenix metropolitan region, USA," *Landscape Urban Plann.* **105**, 149–159 (2012).

19. R. de Jong et al., "Analysis of monotonic greening and browning trends from global NDVI time-series," *Remote Sens. Environ.* **115**(2), 692–702 (2011).

20. S. J. Jeong et al., "Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008," *Glob. Change Biol.* **17**, 2385–2399 (2011).

21. J. Pastor-Guzman, J. Dash, and P. M. Atkinson, "Remote sensing of mangrove forest phenology and its environmental drivers," *Remote Sens. Environ.* **205**, 71–84 (2018).

22. S. Piao et al., "Altitude and temperature dependence of change in the spring vegetation green-up date from 1982 to 2006 in the Qinghai-Xizang Plateau," *Agric. For. Meteorol.* **151**(12), 1599–1608 (2011).

23. M. I. C. Ciocîrlan, A. L. Curtu, and G. R. Radu, "Predicting leaf phenology in forest tree species using UAVs and satellite images: a case study for European Beech (*Fagus sylvatica* L.)," *Remote Sens.* **14**, 6198 (2022).

24. D.-L. Zhai and J.-C. Xu, "The legacy effects of rubber defoliation period on the refoliation phenology, leaf disease, and latex yield," *Plant Diversity* **45**, 98–103 (2023).

25. L. Gasparotto and J. C. R. Pereira, *Doenças da seringueira no Brasil*, EMBRAPA, Brasília (2012).

26. BMKG, Meteorological, Climatological, and Geophysical Agency website, https://dataonline.bmkg.go.id/ (accessed 20 March 2023).

27. S. Chen et al., "Cloud removal with SAR-optical data fusion and graph-based feature aggregation network," *Remote Sens.* **14**, 3374 (2022).

28. C. H. Lin et al., "Radiometric normalization and cloud detection of optical satellite images using invariant pixels," *ISPRS J. Photogramm. Remote Sens.* **106**, 107–117 (2015).

29. X. H. Li et al., "Monitoring snow cover variability (2000–2014) in the Hengduan Mountains based on cloud-removed MODIS products with an adaptive spatio-temporal weighted method," *J. Hydrol.* **551**, 314–327 (2017).

30. H. Zhai et al., "Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.* **144**, 235–253 (2018).

31. X. Zhu and E. Helmer, "An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions," *Remote Sens. Environ.* **214**, 135–153 (2018).

32. A. H. Sancheze et al., "Comparison of cloud cover detection algorithms on Sentinel–2 images of the Amazon tropical forest," *Remote Sens.* **12**, 1284 (2020).

33. T. Adão et al., "Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sens.* **9**, 1110 (2017).

34. F. Mouret et al., "Reconstruction of Sentinel-2 derived time series using robust Gaussian mixture models: application to the detection of anomalous crop development," *Comput. Electron. Agric.* **198**, 106983 (2022).

35. C. M. Gevaert and F. J. García-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.* **156**, 34–44 (2015).

36. C. Senf et al., "Mapping land cover in complex Mediterranean landscapes using Landsat: improved classification accuracies from integrating multi-seasonal and synthetic imagery," *Remote Sens. Environ.* **156**, 527–536 (2015).

37. Y. Zhang et al., "Learning-based spatial–temporal superresolution mapping of forest cover with MODIS images," *IEEE Trans. Geosci. Remote Sens.* **55**, 600–614 (2017).

38. J. Inglada et al., "Operational high resolution land cover map production at the country scale using satellite image time series," *Remote Sens.* **9**, 95 (2017).

39. P. Griffiths et al., "A pixel-based Landsat compositing algorithm for large area land cover mapping," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **6**(5), 2088–2101 (2012).

40. T. Hermosilla et al., "Disturbance-informed annual land cover classification maps of Canada's forested ecosystems for a 29-year Landsat time series," *Can. J. Remote Sens.* **44**, 67–87 (2018).

41. S. C. Chapra and R. P. Canale, *Numerical Methods for Engineers*, McGraw-Hill, Singapore (1998).

42. Z. Cai et al., "Performance of smoothing methods for reconstructing NDVI time series and estimating vegetation phenology from MODIS data," *Remote Sens.* **9**(12), 1271 (2017).

43. P. M. Atkinson et al., "Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology," *Remote Sens. Environ.* **123**, 400–417 (2012).

44. S. Li et al., "High-quality vegetation index product generation: a review of NDVI time series reconstruction techniques," *Int. J. Appl. Earth Observ. Geoinf.* **105**, 102640 (2021).
45. P. Seneviratne et al., "Effect of tapping rest during wintering months on latex yield and tapping panel dryness in three rubber (*Hevea brassiliensis*) clones," *J. Rubber Res. Inst. Sri Lanka* **100**, 1–21 (2020).
46. L. Breiman, "Bagging predictors," *Mach. Learn.* **24**(2), 123–140 (1996).
47. R. Das et al., "Factors affecting sporulation of *Pestalotiopsis disseminata* causing grey blight disease of *Persea bombycina* Kost., the primary food plant of muga silkworm," *Crop Protect.* **29**, 963–968 (2010).
48. A. K. H. Priyashantha et al., "Plant–fungi interactions: where it goes?" *Biology* **12**, 809 (2023).
49. A. Wongcharoen, "Biological activities of soils under rubber trees (*Hevea brasiliensis*) and interactions with trunk phloem necrosis," Biodiversity. Université Pierre et Marie Curie - Paris VI (2010). English. ffNNT : 2010PA066252ff.
50. L. Wang, "Physiological and molecular responses to drought stress in rubber tree (*Hevea brasiliensis* Muell. Arg.)," *Plant Physiol. Biochem.* **83**, 243–249 (2014).
51. A. N. Cahyo et al., "Rubber genotypes with contrasting drought factor index revealed different mechanisms for drought resistance in *Hevea brasiliensis*," *Plants* **11**, 3563 (2022).

Yeni Herdiyeni received her Doctor of Computer Science degree from the University of Indonesia in 2010. She is currently an associate professor in the Computer Science Department, Faculty of Mathematics and Natural Sciences at IPB University. Her expertise mainly includes computer vision, remote sensing, computational intelligence, advanced and big data analytics, and text mining. She is currently involved in a project for applying remote sensing for biodiversity monitoring.

Biographies of the other authors are not available.