Use of spectral reflectance from a compact spectrometer to assess chlorophyll content in Zizania latifolia

Use of spectral reflectance from a compact spectrometer to assess chlorophyll content in *Zizania latifolia*

Rei Sonobe

Faculty of Agriculture, Shizuoka University, Shizuoka, Japan

Hiroto Yamashita

Faculty of Agriculture, Shizuoka University, Shizuoka, Japan

United Graduate School of Agricultural Science, Gifu University, Gifu, Japan

Haruyuki Seki, Akio Morita and Takashi Ikka

Faculty of Agriculture, Shizuoka University, Shizuoka, Japan

Corresponding author

Rei sonobe (E–mail: sonobe.rei@shizuoka.ac.jp)

Use of spectral reflectance from a compact spectrometer to assess chlorophyll content in *Zizania latifolia*

Hyperspectral remote sensing is frequently used to monitor chlorophyll content, an important characteristic for assessing photosynthetic ability, health and defence against a variety of degenerative diseases. To obtain hyperspectral data, field portable spectroradiometers, such as Ocean Optics Hyperspectral Vis-NIR spectroradiometers and Analytical Spectral Devices FieldSpec series, have been widely used. However, the prices of these devices are above consumer levels, which prevents practical use. The development of an affordable hyperspectral remote sensing system would be advantageous. Highly sensitive, affordable and finger-tip size spectrometers have recently been released. In this study we investigate the potential of hyperspectral data obtained from such a compact spectrometer (C12880MA-10, Hamamatsu Photonics) for estimating chlorophyll content in *Zizania latifolia*. We also tested the efficacy of five pre-processing techniques (first derivative reflectance, continuum-removal transformation, de-trending, multiplicative scatter correction and standard normal variate) in conjunction with five machine learning algorithms.

Keywords: C12880MA-10; Cubist; hyperspectral; sensitivity analysis

1. Introduction

The close relationship between primary production and chlorophyll content is well-known. Of all biochemical variables, chlorophyll content is a most important indicator of photosynthetic activity (Peng *et al.*, 2011). Furthermore, chlorophyll content has been used for evaluating various plant stresses since changes in chlorophyll content of leaves are related to the effects of disease as well as nutritional and environmental stresses (Zhu *et al.*, 2019). Estimating chlorophyll content is therefore useful for monitoring growth rates and managing fertilizing schedules.

Although spectrophotometric measurements using ultraviolet and visible (UV-VIS) spectroscopy or high-performance liquid chromatography (HPLC) measurements have

been widely adopted to accurately measure chlorophyll content (Prado-Cabrero *et al.*, 2016), these techniques are expensive, labour-intensive and require bulky equipment (Kalaji *et al.*, 2017). Portable equipment, such as the SPAD-502 Leaf Chlorophyll Meter (Konica Minolta), provide a simpler method for quantifying chlorophyll by measuring leaf absorbance at specific wavelengths. However, the light transmittance of a leaf is also influenced by leaf structure, water content and leaf pigment distribution (Padilla *et al.*, 2018). Portable chlorophyll meters were shown to be unsuitable for quantifying chlorophyll content in Manchurian wild rice (*Zizania latifolia*) since it is silica-rich, and silica content is correlated with leaf structure (Kindomihou *et al.*, 2006).

Hyperspectral remote sensing, which provides spectral information on continuous wavelengths, has played an important role in evaluating vegetation characteristics and is a technique that assists in management of agricultural fields (Huang *et al.*, 2016). Chlorophyll absorbs energy strongly in the ultraviolet, blue and red regions, resulting in weak reflectance and transmittance (Roy, 1989). Reflectance from vegetation is thus low in the blue (400 to 500 nm) and red (650 to 690 nm) spectral regions (Datt, 1999). Wavelengths at the green peak (540 to 560 nm) and the red edge (680 to 750 nm) in particular have been used for estimating chlorophyll content (Carter and Knapp, 2001, Zarco-Tejada *et al.*, 2001).

Field portable spectroradiometers are expensive and there are practical difficulties in using them, thus development of an affordable hyperspectral remote sensing system would be advantageous (Uto *et al.*, 2016). Recently, affordable fingertip-sized spectrometers for obtaining hyperspectral data have been developed, which may be attached to unmanned aerial vehicles or used for sensing over a large area (Chen *et al.*, 2018). In the present study, a compact spectrometer (C12880MA-10, Hamamatsu Photonics) was used to evaluate chlorophyll content in *Zizania latifolia*.

Pre-processing techniques are used to remove physical phenomena in the spectra in order to improve subsequent multivariate regression, classification model or exploratory analysis (Rinnan *et al.*, 2009). A first derivative reflectance (FDR) or continuum-removal (CR) transformation may be run to emphasise spectral features, such as the red edge and green peak (Demetriadesshah *et al.*, 1990, Schmidt and Skidmore, 2003). De-trending (DT), multiplicative scatter correction (MSC) and standard normal variate (SNV) are also known to reduce noise or baseline shift caused by light scattering in raw reflectance data (Barnes *et al.*, 1989, Liang *et al.*, 2020). However, the advantages of some techniques may be obscured depending on the characteristics of objective variables, such as chemical variations (Wu *et al.*, 2019). We therefore tested these techniques and evaluated their potential for improving estimation accuracies.

The benefits of machine learning algorithms for identifying vegetation characteristics have also been demonstrated. Cubist-based regression, stochastic gradient boosting (SGB; (Fernandez-Delgado *et al.*, 2019, Breunig *et al.*, 2020), deep belief nets (DBN;(Sonobe *et al.*, 2020a), kernel-based extreme learning machines (KELM; (Sonobe *et al.*, 2018b), random forests (RF; (Cui *et al.*, 2019) and support vector machines (SVM; (Yang *et al.*, 2011, Wang *et al.*, 2013) performed well during studies assessing biochemical or physiological features of vegetation. A challenge of applying machine learning algorithms is optimizing their hyperparameters and selecting a combination of hyperparameters that corresponds to a convex optimization problem (Abedi *et al.*, 2012). Although grid search strategies have been adopted to solve this problem (Puertas *et al.*, 2013), these strategies are inadequate for configuring algorithms for new data sets, and a random search is considered a superior approach (Bergstra and Bengio, 2012). Furthermore, Bayesian optimization, which is a framework used to optimize hyperparameters of noisy, expansive black-box functions,

constitutes a structured approach to modelling uncertainty and may offer better selections (Xia *et al.*, 2017). We therefore used Bayesian optimization to optimize the hyperparameters of machine learning algorithms.

The main objectives of this study were (1) to evaluate the potential of hyperspectral data derived from a C12880MA-10 device for estimating chlorophyll content in *Zizania latifolia*, and (2) to identify which combination of pre-processing technique and machine learning algorithms is the most suitable for constructing chlorophyll content estimation models.

2. Materials and methods

2.1. Measurements and datasets

Manchurian wild rice (*Zizania latifolia*) plants were cultivated in flooded conditions with inter-row and within-row spacing 100 cm apart, on a paddy field at Shizuoka University, Shizuoka, Japan (Figure 1). A total of 200 leaves were measured for reflectance and chlorophyll content on 2 and 5 October, 2020.

A complementary metal-oxide-semiconductor (CMOS) sensor (C12880MA-10, Hamamatsu Photonics, Table 1) and a shape memory alloy (SMA) to SMA fiber patch cable (M25L05, Thorlabs, Inc.) with numerical aperture 0.22, were assembled and hyperspectral reflectance was measured at nadir, 3 cm above the samples. The grating equations provided by Hamamatsu Photonics were used to convert pixel number to wavelength and the data were resampled in 5-nm bands across the wavelength domain from 340 to 850 nm.

<Figure 1>

<Table 1>

Reflectance of the target was calculated using the following equation:

$$\rho_\lambda = \frac{S_\lambda - D_\lambda}{W_\lambda - D_\lambda}, \qquad (1)$$

where S, W and D are the target, a diffuse reflectance standard (WS-1, Ocean Optics) and dark current at wavelength λ nm.

A dual-beam scanning ultra violet-visible spectrophotometer (UV–1900, Shimadzu, Japan), dimethylformamide extraction and Porra's method (Porra *et al.*, 1989) were used to obtain real chlorophyll content values.

2.2.Spectral pre-processing

The combinations of software and packages used for applying various processes are listed in Table 2. Spectral pre-processing techniques were used to reduce noise and insignificant signals in the spectra and to improve the subsequent multivariate regression, classification model or exploratory analysis. They were mainly divided to two groups: scatter-correction methods and spectral derivatives (Rinnan *et al.*, 2009). De-trending (DT), multiplicative scatter correction (MSC) and standard normal variate (SNV) belong in the first group and first derivative reflectance (FDR) belongs in the latter. Further, continuum removal (CR), which is a brightness normalization technique that models water stress comparatively well (Boloorani et al. 2020), was also applied to enhance associated changes (Clark and Roush, 1984). These five pre-processing techniques were evaluated together with the original reflectance (OR) from the compact spectrometer.

<Table 2>

2.3. Regression model generation based on machine learning algorithms

Cubist is a rule-based model tree method; a multivariate linear regression model is used for fitting at each leaf node on the tree (Quinlan, 1992). Committee models and instance-based corrections using nearest neighbours were also used to improve predictive accuracy. Cubist has the ability to deal with nonlinear and complex relationships between dependent and independent variables using both continuous and categorical input variables (Chen *et al.*, 2020). Further, Cubist is known to select spectral variables efficiently (Sonobe *et al.*, 2020b).

Deep belief nets (DBN) consist of stacked modules of restricted Boltzmann machines (RBMs), which is an undirected energy-based model with two layers of visible and hidden units. Each RBM module is trained individually in an unsupervised manner using a contrastive divergence procedure (Hinton *et al.*, 2006). Dropout is used as an input of the subsequent RBM stage during the training phase to facilitate high-quality predictions, and the whole network is commonly trained using a supervised learning approach called the fine-tuning method.

Kernel-based extreme learning machine, which is expressed as a single hidden layer feed-forward neural network, has been widely applied with an RBF kernel for many practical tasks, such as prediction, fault diagnosis, recognition, classification and signal processing (Li *et al.*, 2016, Sonobe, 2019a, Sonobe, 2019b). The method has been further developed, and improvements have increased the number of hyperparameters included, thereby reducing the advantages gained from limiting their number. In the present study, we used KELM as proposed by (Huang *et al.*, 2012).

Random forests (RF) is an ensemble learning algorithm composed of many decision tree models. These trees are combined to provide more accurate prediction results

(Breiman, 2001). It performs well for both classification (Caglayan *et al.*, 2020, Mansaray *et al.*, 2020a) and regression(Mansaray *et al.*, 2020b).

Stochastic gradient boosting (SGB) has both boosting and bagging advantages as well as the ability to model nonlinear relationships, manage qualitative and quantitative variables and remain robust despite missing values and data outliers (Friedman, 2002). Furthermore, its limited number of hyperparameters (i.e. total number of trees to fit, maximum depth of each tree, learning rate and minimum number of observations in the terminal nodes of the trees) makes it effective for generating robust models (Greenwell *et al.*, 2020). However, a trade-off between the number of trees and learning rate has been identified (Friedman, 2002).

Support vector machines (SVM) have been used to solve non-linear problems by mapping the input variables into higher dimensional feature space using a Gaussian radial basis function (RBF) kernel(Al-Fugara *et al.*, 2020). However, it is necessary to optimize the hyperparameters to avoid overfitting (Reda *et al.*, 2020). There are two hyperparameters: C, which controls the trade-off between the smooth decision limit and the regression training points, and gamma, which defines how far the influence of a single training sample reaches.

2.4.Performance assessment

The performance of each algorithms was evaluated based on the ratio of performance to deviation (RPD, Equation (2); (Williams and Norris, 1987). (Chang *et al.*, 2001) grouped results into three categories according to RPD values: ‘A’ (RPD > 2.0), ‘B’ ($1.4 \leq \text{RPD} \leq 2.0$) and ‘C’ (RPD < 1.4). Regression models classified into ‘A’ or ‘B’ were assumed to have the potential to estimate chlorophyll content.

$$\text{RPD} = \text{SD}/\text{RMSE} \qquad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(\widehat{y_i} - y_i)^2} \quad (3)$$

where SD is the standard deviation of chlorophyll content in the test data, RMSE is root-mean-square error, n is number of samples, y_i is measured chlorophyll content and $\widehat{y_i}$ is estimated chlorophyll content.

3. Results

3.1.Chlorophyll content and composition

The chlorophyll content per leaf area (cm²) ranged from 17.53 to 58.0 µg for total chlorophyll, 14.47 to 45.84 µg for chlorophyll-a and 3.06 to 12.19 µg for chlorophyll-b (Figure 2). The chlorophyll-a to chlorophyll-b ratio ranged from 3.47 to 5.24 (Figure 2).

<Figure 2>

3.2. Spectral patterns after pre-processing

Figure 3 shows the correlation coefficients between chlorophyll content and original reflectance of the pre-processed spectra. A significant positive correlation ($r=0.641$, $p < 0.001$) was confirmed at 585 nm for FDR. Generally, however, the pre-processing techniques weakened the correlations. The lowest negative correlation coefficient was -0.875 for OR. The wavelengths showing the strongest negative correlation coefficients varied along the spectrum: near the green peak (520 nm) for OR, and near the red edge for FDR, CR, DT, MSC and SNV (690 (r=-0.750), 730 (r=-0.379), 710 (r=-0.350), 715 nm (r=-0.335) and 715 nm (r=0.349), respectively).

3.3.Accuracy validation

The chlorophyll content estimation accuracy for each combination of pre-processing technique and machine learning algorithm was calculated from 100 repetitions (i.e. 25

measurements × 100 repetitions). The best pre-processing and algorithm combination were OR and Cubist with an RPD of 2.01 (in the 'A' category) and an RMSE of 3.80 μg cm^2 (Tables 3 and 4). OR performed satisfactorily with the other machine learning algorithms except for SVM. FDR also performed adequately with Cubist, KELM and RF. However, the scatter-correction methods and CR were not effective for enhancing the spectral features related to chlorophyll content from reflectance data from a C12880MA-10 device.

Among the 100 repetitions, the lowest RPD value was 1.4 for the OR-Cubist combination, implying that it was always suitable for estimating chlorophyll content in *Zizania latifolia* using the compact spectrometer. RPDs of the other combinations, however, were below 1.1.

<Table 3>

<Table 4>

3.4.Sensitivity analysis

The importance of each wavelength of OR was evaluated with data-based sensitivity analysis (DSA) for each algorithm at 20 nm intervals (Figure 4). The highest importance was identified near the green peak (540-560 nm for Cubist, DBN, KELM and SVM, 570 nm for RF and 520-540 nm for SGB) and the highest values were 11.3, 8.99, 8.07, 10.16, 14.16 and 7.62% for Cubist, DBN, KELM, RF, SGB and SVM, respectively. Smaller peaks were also confirmed at the start of the red edge near 630 nm (5.21, 5.89, 5.49, 6.71 and 8.39% for Cubist, DBN, KELM, RF and SGB, respectively); however, this was obscure for SVM (4.55% at 620-640 nm). Importance was almost zero for SGB (0.76%) at wavelengths greater than 720 nm but was still present for other algorithms (10.74, 10.81, 14.87, 7.00 and 19.16% for Cubist, DBN, KELM, RF, SGB and SVM, respectively). Above 760 nm, however, importance values were below 10%,

with the exception of KELM and SVM (10.52 and 13.81%, respectively).

<Figure 4>

4. Discussion

4.1.Spectral features of a compact spectrometer

For the compact spectrometer, C12880MA, outputs below 400 nm were low and noisy since at that wavelength sensitivity was low and sunlight illumination poor (Uto *et al.*, 2016). During our investigation of the C12880MA-10, we also recorded low reflectance values from our samples, but noise below the wavelength of 400 nm was negligible. In our study, the distance between the tip of the fiber and the target was only 3 cm, which might reduce diffuse light and background effects. However, imaging spectroscopy through combining the potential of digital images with hyperspectral measurements may benefit the functioning of agricultural systems and rangelands (Heiden *et al.*, 2016). It would be beneficial to assess the C12880MA-10 hyperspectral sensor with whisk broom scanning.

However, reflectance at 750 nm was greater than 0.60, which is the highest reflectance value at 750 nm reported in the online dataset (ANGERS, for Populus alba L.; (Feret *et al.*, 2008). Furthermore, a decrease at 800 nm was identified and there were no clear trends. Relative sensitivity was less than 0.5 at 700 nm (HamamatsuPhotonics, 2019); it would thus be inappropriate to use reflectance from the compact spectrometer to estimate chlorophyll content at wavelengths greater than 700 nm. Nevertheless, the red edge and related indicators such as the red-edge inflection point have been used for evaluating chlorophyll content in previous studies (Miller *et al.*, 1990, Flynn *et al.*, 2020, Sharifi, 2020).

4.2.Effects of pre-processing techniques on reflectance from a compact spectrometer

The combination of OR and Cubist performed best, indicating that large spectral modifications are not necessary for original reflectance obtained from a C12880MA-10 device. In fact, scatter-correction methods and CR reduced OR abilities. Although scatter-correction methods have been shown to perform poorly when the raw spectra data include large chemical variations (Wu *et al.*, 2019), the standard deviation of chlorophyll content in *Zizania latifolia* (7.66 μg cm^2) is small compared to that of other species (such as wasabi [7.86 μg cm^2;] and tea [26.50 μg cm^2;]) in which DT and certain forms of MSC and SNV perform well. However, the devices used to acquire these reflectance data were different. The Analytical Spectral Device (ASD) FieldSpec or ImSpector N10E high-spectrometer, which has a narrower full width at half maximum, was used in some studies that reported a high performance of DT, MSC and SNV (Yu *et al.*, 2014, Zhang *et al.*, 2016, Golhani *et al.*, 2019, Sonobe *et al.*, 2020c, Yamashita *et al.*, 2020). Furthermore, reflectance data intervals from C12880MA-10 might be relatively coarse in wavelength. Moreover, irregular reflectance data resulted in more continuum points (Lehnert, 2020), and CR failed to enhance the spectral features related to chlorophyll absorption, since a spectral curve with values between 0 and 1 was obtained by normalizing OR (Clark and Roush, 1984).

4.3.Performance of different machine learning algorithms

Cubist, a regression model, has been shown to perform well on various dataset, but it occasionally fails when used on large datasets (Fernandez-Delgado *et al.*, 2019). Some studies found that KELM was superior to Cubist when using fine hyperspectral reflectance data such as reflectance from an ASD FieldSpec (Sonobe *et al.*, 2020c, Sonobe *et al.*, 2020b). Reflectance from the red edge domain is usually important for an

accurate estimation of chlorophyll content. Normally, importance is dispersed for KELM or SVM and a degree of importance over the non-informative domain was shown (Sonobe *et al.*, 2018a). However, the sensitivity of C12880MA-10 in this domain was poor, and removing that influence might be effective. Since Cubists importance in this domain was lower than that of KELM and SVM, Cubist was less influenced by the noise in the domain. In contrast, RF and SGB were less accurate, although their importance at 700 nm was lower than Cubist's. Unlike Cubist, SGB performs well with large and difficult datasets (Fernandez-Delgado *et al.*, 2019), and thus failed in our study probably due to the small size of the training data set. Using only a fraction of the training data for SGB may have led to overfitting. When running RF, similar strategies produced a less accurate estimation, since one third of the training data is separated as out-of-bag (OOB) samples; these samples are not used for training the tree but to evaluate performance. This strategy may have reduced the sample size too much to generate regression models.

5. Conclusions

We examined the relationship between *Zizania latifolia* chlorophyll content and reflectance measured using a compact spectrometer, C12880MA-10 (Hamamatsu Photonics). To analyse reflectance data, we evaluated the efficacies of pre-processing techniques: first derivative reflectance (FDR), continuum-removal (CR) transformation, de-trending (DT), multiplicative scatter correction (MSC) and standard normal variate (SNV), as well as original reflectance (OR). To this end we also used five machine learning algorithms: random forests (RF), support vector machine (SVM), kernel-based extreme learning machine (KELM), Cubist and Stochastic Gradient Boosting (SGB).

OR was selected as the best pre-processing technique when combined with Cubist, indicating that large spectral modifications are not necessary for estimating chlorophyll

content from reflectance data from C12880MA-10. This illustrates the potential of using data from this compact spectrometer. However, the reflectance at the red edge should be excluded from analyses due to its low sensitivity.

The proposed method is cost effective, practical for consumers to apply and will enable effective crop management. To make use of imaging spectroscopy the C12880MA-10 hyperspectral sensor with whisk broom scanning is required in addition; this may be of added benefit to precision agriculture.

Acknowledgement

We thank Mr. Keitaro Koike, Mr. Junji Shinada, Ms. Shiori Yonezawa, and Ms. Ibuki Yamazaki of Shizuoka University for assisting with the plant sampling and chlorophyll measurement.

Funding

This study was supported financially by JSPS KAKENHI (grant number 19K06313).

Disclosure statement

There are no conflicts of interest.

References

Abedi, M., Norouzi, G.H. & Bahroudi, A., 2012. Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences,* 46**,** 272-283.

Al-Fugara, A., Ahmadlou, M., Al-Shabeeb, A.R., Alayyash, S., Al-Amoush, H. & Al-Adamat, R., 2020. Spatial mapping of groundwater springs potentiality using grid search-based and genetic algorithm-based support vector regression. *Geocarto International***,** 20.

Barnes, R.J., Dhanoa, M.S. & Lister, S.J., 1989. Standardnormalvariatetransformationandde-trendingofnear-infrareddiffusereflectancespectra. *Applied Spectroscopy,* 43**,** 772-777.

Bergstra, J. & Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research,* 13**,** 281-305.

Breiman, L., 2001. Random forests. *Machine Learning,* 45**,** 5-32.

Breiman, L., Cutler, A., Liaw, A. & Wiener, M., 2018. *Breiman and Cutler's Random Forests for Classification and Regression* [online]. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf [Accessed Access Date 27 November, 2020].

Breunig, F.M., Galvao, L.S., Dalagnol, R., Dauve, C.E., Parraga, A., Santi, A.L., Della Flora, D.P. & Chen, S.S., 2020. Delineation of management zones in agricultural fields using cover crop biomass estimates from PlanetScope data. *International Journal of Applied Earth Observation and Geoinformation,* 85.

Caglayan, S.D., Leloglu, U.M., Ginzler, C., Psomas, A., Zeydanli, U.S., Bilgin, C.C. & Waser, L.T., 2020. Species level classification of Mediterranean sparse forests-maquis formations using Sentinel-2 imagery. *Geocarto International***,** 20.

Carter, G.A. & Knapp, A.K., 2001. Leaf optical properties in higher plants: Linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany,* 88**,** 677-684.

Chang, C.W., Laird, D.A., Mausbach, M.J. & Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal,* 65**,** 480-490.

Chen, J., Cai, F.H., He, R.X. & He, S.L., 2018. Experimental Demonstration of Remote and Compact Imaging Spectrometer Based on Mobile Devices. *Sensors,* 18**,** 10.

Chen, Q., Yang, X.C., Ouyang, Z.T., Zhao, N.Z., Jiang, Q.T., Ye, T.T., Qi, J. & Yue, W.Z., 2020. Estimation of anthropogenic heat emissions in China using Cubist with points-of-interest and multisource remote sensing data. *Environmental Pollution,* 266.

Clark, R.N. & Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research,* 89**,** 6329-6340.

Cui, L., Jiao, Z.T., Dong, Y.D., Sun, M., Zhang, X.N., Yin, S.Y., Ding, A.X., Chang, Y.X., Guo, J. & Xie, R., 2019. Estimating Forest Canopy Height Using MODIS BRDF Data Emphasizing Typical-Angle Reflectances. *Remote Sensing,* 11**,** 21.

Datt, B., 1999. A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using Eucalyptus leaves. *Journal of Plant Physiology,* 154**,** 30-36.

Demetriadesshah, T.H., Steven, M.D. & Clark, J.A., 1990. High resolution derivative spectra in remote sensing. *Remote Sensing of Environment,* 33**,** 55-64.

Drees, M., Rueckert, J., Hinton, G., Salakhutdinov, R. & Rasmussen, C.E., 2015. *Package for Deep Architectures and Restricted Boltzmann Machines* [online]. https://mran.microsoft.com/snapshot/2016-07-08/web/packages/darch/darch.pdf [Accessed Access Date 27 November, 2020].

Feret, J.B., Francois, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., Le Maire, G. & Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment,* 112**,** 3030-3043.

Fernandez-Delgado, M., Sirsat, M.S., Cernadas, E., Alawadi, S., Barro, S. & Febrero-Bande, M., 2019. An extensive experimental survey of regression methods. *Neural Networks,* 111**,** 11-34.

Flynn, K.C., Frazier, A.E. & Admas, S., 2020. Performance of chlorophyll prediction indices forEragrostis tefat Sentinel-2 MSI and Landsat-8 OLI spectral resolutions. *Precision Agriculture,* 21**,** 1057-1071.

Friedman, J.H., 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis,* 38**,** 367-378.

Golhani, K., Balasundram, S.K., Vadamalai, G. & Pradhan, B., 2019. Estimating chlorophyll content at leaf scale in viroid-inoculated oil palm seedlings (Elaeis guineensis Jacq.) using reflectance spectra (400 nm-1050 nm). *International Journal of Remote Sensing,* 40**,** 7647-7662.

Greenwell, B., Boehmke, B. & Cunningham, J., 2020. *Generalized Boosted Regression Models* [online]. https://cran.r-project.org/web/packages/gbm/gbm.pdf [Accessed Access Date 27 November, 2020].

Hamamatsuphotonics, 2019. *Mini-spectrometer* [online]. http://www.farnell.com/datasheets/2822646.pdf [Accessed Access Date 27 November, 2020].

Heiden, U., Iwasaki, A., Muller, A., Schlerf, M., Udelhoven, T., Uto, K., Yokoya, N. & Chanussot, J., 2016. Foreword to the Special Issue on Hyperspectral Remote Sensing and Imaging Spectroscopy. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 9**,** 3904-3908.

Hinton, G.E., Osindero, S. & Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. *Neural Computation,* 18**,** 1527-1554.

Huang, G.B., Zhou, H.M., Ding, X.J. & Zhang, R., 2012. Extreme Learning Machine for Regression and Multiclass Classification. *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics,* 42**,** 513-529.

Huang, Y.B., Lee, M.A., Thomson, S.J. & Reddy, K.N., 2016. Ground-based hyperspectral remote sensing for weed management in crop production. *International Journal of Agricultural and Biological Engineering,* 9**,** 98-109.

Kalaji, H.M., Dabrowski, P., Cetner, M.D., Samborska, I.A., Lukasik, I., Brestic, M., Zivcak, M., Tomasz, H., Mojski, J., Kociel, H. & Panchal, B.M., 2017. A comparison between different chlorophyll content meters under nutrient deficiency conditions. *Journal of Plant Nutrition,* 40**,** 1024-1034.

Karatzoglou, A., Smola, A., Hornik, K., Maniscalco, M.A. & Teo, C.H., 2019. *Kernel-Based Machine Learning Lab* [online]. https://cran.r-project.org/web/packages/kernlab/kernlab.pdf [Accessed Access Date 27 November, 2020].

Kindomihou, V., Sinsin, B. & Meerts, P., 2006. Effect of defoliation on silica accumulation in five tropical fodder grass species in Benin. *Belgian Journal of Botany,* 139**,** 87-102.

Kucheryavskiy, S., 2020. *Multivariate Data Analysis for Chemometrics* [online]. https://cran.r-project.org/web/packages/mdatools/mdatools.pdf [Accessed Access Date 27 November, 2020].

Kuhn, M., 2020. *Rule- And Instance-Based Regression Modeling* [online]. https://cran.r-project.org/web/packages/Cubist/Cubist.pdf [Accessed Access Date 27 November, 2020].

Lehnert, L.W., 2020. *Manage, Analyse and Simulate Hyperspectral Data* [online]. https://cran.r-project.org/web/packages/hsdar/hsdar.pdf [Accessed Access Date 27 November, 2020].

Li, X.D., Mao, W.J. & Jiang, W., 2016. Multiple-kernel-learning-based extreme learning machine for classification design. *Neural Computing & Applications,* 27**,** 175-184.

Liang, K., Huang, J.N., He, R.Y., Wang, Q.J., Chai, Y.Y. & Shen, M.X., 2020. Comparison of Vis-NIR and SWIR hyperspectral imaging for the non-

destructive detection of DON levels in Fusarium head blight wheat kernels and wheat flour. *Infrared Physics & Technology,* 106**,** 9.

Mansaray, L.R., Wang, F.M., Huang, J.F., Yang, L.B. & Kanu, A.S., 2020a. Accuracies of support vector machine and random forest in rice mapping with Sentinel-1A, Landsat-8 and Sentinel-2A datasets. *Geocarto International,* 35**,** 1088-1108.

Mansaray, L.R., Wang, F.M., Kanu, A.S. & Yang, L.B., 2020b. Evaluating Sentinel-1A datasets for rice leaf area index estimation based on machine learning regression models. *Geocarto International***,** 12.

Miller, J.R., Hare, E.W. & Wu, J., 1990. Quantitative characterisation of the red edge reflectance 1. An inverted-Gaussian model. *International Journal of Remote Sensing,* 11**,** 1755-1773.

Padilla, F.M., Gallardo, M., Pena-Fleitas, M.T., De Souza, R. & Thompson, R.B., 2018. Proximal Optical Sensors for Nitrogen Management of Vegetable Crops: A Review. *Sensors,* 18.

Peng, Y., Gitelson, A.A., Keydan, G., Rundquist, D.C. & Moses, W., 2011. Remote estimation of gross primary production in maize and support for a new paradigm based on total crop chlorophyll content. *Remote Sensing of Environment,* 115**,** 978-989.

Porra, R.J., Thompson, W.A. & Kriedemann, P.E., 1989. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 975**,** 384-394.

Prado-Cabrero, A., Beatty, S., Howard, A., Stack, J., Bettin, P. & Nolan, J.M., 2016. Assessment of lutein, zeaxanthin and meso-zeaxanthin concentrations in dietary supplements by chiral high-performance liquid chromatography. *European Food Research and Technology,* 242**,** 599-608.

Puertas, O.L., Brenning, A. & Meza, F.J., 2013. Balancing misclassification errors of land cover classification maps using support vector machines and Landsat imagery in the Maipo river basin (Central Chile, 1975-2010). *Remote Sensing of Environment,* 137**,** 112-123.

Quinlan, J.R., Year. Learning with Continuous Classe.ed.^eds. *5th Australian Joint Conference on Artificial Intelligence,* Hobart, TAS, Australia, 343-348.

Reda, R., Saffaj, T., Itqiq, S.E., Bouzida, I., Saidi, O., Yaakoubi, K., Lakssir, B., El Mernissi, N. & El Hadrami, E., 2020. Predicting soil phosphorus and studying the effect of texture on the prediction accuracy using machine learning combined with near-infrared spectroscopy. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy,* 242**,** 11.

Rinnan, A., Van Den Berg, F. & Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trac-Trends in Analytical Chemistry,* 28**,** 1201-1222.

Roy, P.S., 1989. Spectralreflectancecharacteristicsofvegetationandtheiruseinestimatingproductive potential. *Proceedings of the Indian Academy of Sciences-Plant Sciences,* 99**,** 59-81.

Schmidt, K.S. & Skidmore, A.K., 2003. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment,* 85**,** 92-108.

Sharifi, A., 2020. Remotely sensed vegetation indices for crop nutrition mapping. *Journal of the Science of Food and Agriculture,* 100**,** 5191-5196.

Sonobe, R., 2019a. Combining ASNARO-2 XSAR HH and Sentinel-1 C-SAR VH/VV Polarization Data for Improved Crop Mapping. *Remote Sensing,* 11.

Sonobe, R., 2019b. Parcel-Based Crop Classification Using Multi-Temporal TerraSAR-X Dual Polarimetric Data. *Remote Sensing,* 11.

Sonobe, R., Hirono, Y. & Oi, A., 2020a. Quantifying chlorophyll-aandbcontent in tea leaves using hyperspectral reflectance and deep learning. *Remote Sensing Letters,* 11**,** 933-942.

Sonobe, R., Miura, Y., Sano, T. & Horie, H., 2018a. Monitoring Photosynthetic Pigments of Shade-Grown Tea from Hyperspectral Reflectance. *Canadian Journal of Remote Sensing,* 44**,** 104-112.

Sonobe, R., Sano, T. & Horie, H., 2018b. Using spectral reflectance to estimate leaf chlorophyll content of tea with shading treatments. *Biosystems Engineering,* 175**,** 168-182.

Sonobe, R., Yamashita, H., Mihara, H., Morita, A. & Ikka, T., 2020b. Estimation of Leaf Chlorophyll a, b and Carotenoid Contents and Their Ratios Using Hyperspectral Reflectance. *Remote Sensing,* 12**,** 3265.

Sonobe, R., Yamashita, H., Mihara, H., Morita, A. & Ikka, T., 2020c. Hyperspectral reflectance sensing for quantifying leaf chlorophyll content in wasabi leaves using spectral pre-processing techniques and machine learning algorithms. *International Journal of Remote Sensing*.

Stevens, A. & Ramirez-Lopez, L., 2020. *Package 'prospectr* [online]. https://cran.r-project.org/web/packages/prospectr/prospectr.pdf [Accessed Access Date 27 November, 2020].

Uto, K., Seki, H., Saito, G., Kosugi, Y. & Komatsu, T., 2016. Development of a Low-Cost Hyperspectral Whiskbroom Imager Using an Optical Fiber Bundle, a Swing Mirror, and Compact Spectrometers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 9**,** 3909-3925.

Wang, F.M., Huang, J.F., Wang, Y., Liu, Z.Y. & Zhang, F.Y., 2013. Estimating nitrogen concentration in rape from hyperspectral data at canopy level using support vector machines. *Precision Agriculture,* 14**,** 172-183.

Williams, P. & Norris, K., 1987. *Near- Infrared Technology in the Agricultural and Food Industries* St. Paul, Minnesota: American Association of Cereal Chemists Inc.

Wu, Y.F., Peng, S.L., Xie, Q., Han, Q.J., Zhang, G.W. & Sun, H.G., 2019. An improved weighted multiplicative scatter correction algorithm with the use of variable selection: Application to near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems,* 185**,** 114-121.

Xia, Y.F., Liu, C.Z., Li, Y.Y. & Liu, N.N., 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications,* 78**,** 225-241.

Yamashita, H., Sonobe, R., Hirono, Y., Morita, A. & Ikka, T., 2020. Dissection of hyperspectral reflectance to estimate nitrogen and chlorophyll contents in tea leaves based on machine learning algorithms. *Scientific Reports,* 10**,** 17360.

Yang, X.H., Huang, J.F., Wu, Y.P., Wang, J.W., Wang, P., Wang, X.M. & Huete, A.R., 2011. Estimating biophysical parameters of rice with remote sensing data using support vector machines. *Science China-Life Sciences,* 54**,** 272-281.

Yu, K., Lenz-Wiedemann, V., Chen, X.P. & Bareth, G., 2014. Estimating leaf chlorophyll of barley at different growth stages using spectral indices to reduce soil background and canopy structure effects. *Isprs Journal of Photogrammetry and Remote Sensing,* 97**,** 58-77.

Zarco-Tejada, P.J., Miller, J.R., Noland, T.L., Mohammed, G.H. & Sampson, P.H., 2001. Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing,* 39**,** 1491-1507.

Zhang, J.F., Han, W.T., Huang, L.W., Zhang, Z.Y., Ma, Y.M. & Hu, Y.M., 2016. Leaf Chlorophyll Content Estimation of Winter Wheat Based on Visible and Near-Infrared Sensors. *Sensors,* 16**,** 11.

Zhu, M.Y., Yang, H.B. & Li, Z.W., 2019. Early Detection and Identification of Rice Sheath Blight Disease Based on Hyperspectral Image and Chlorophyll Content. *Spectroscopy and Spectral Analysis,* 39**,** 1898-1904.

Lists of the Figures

Lists of the Tables

Figure 1. *Zizania latifolia* cultivated in a paddy field.

Figure 2. Histograms of (a) total chlorophyll, (b) chlorophyll-a, (c) chlorophyll-b contents and (d) chlorophyll-a to chlorophyll-b ratio.

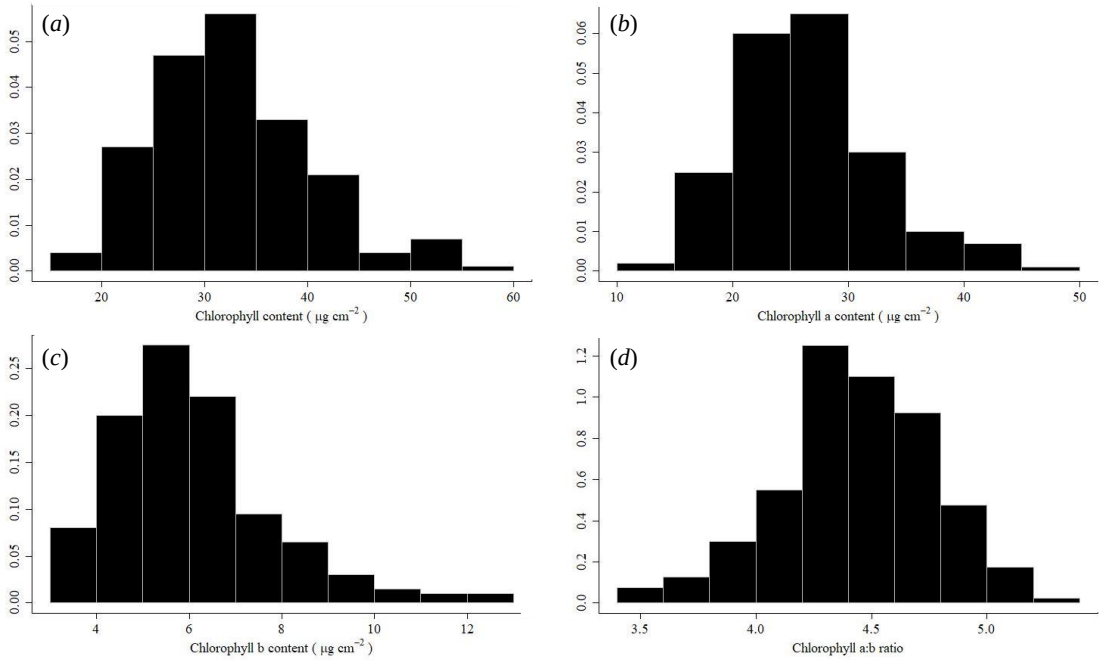

Figure 3. Correlations between chlorophyll content and original reflectance (OR) or a pre-processed spectrum: first derivative reflectance (FDR), continuum-removal (CR) transformation, de-trending (DT), multiplicative scatter correction (MSC) and standard normal variate (SNV).

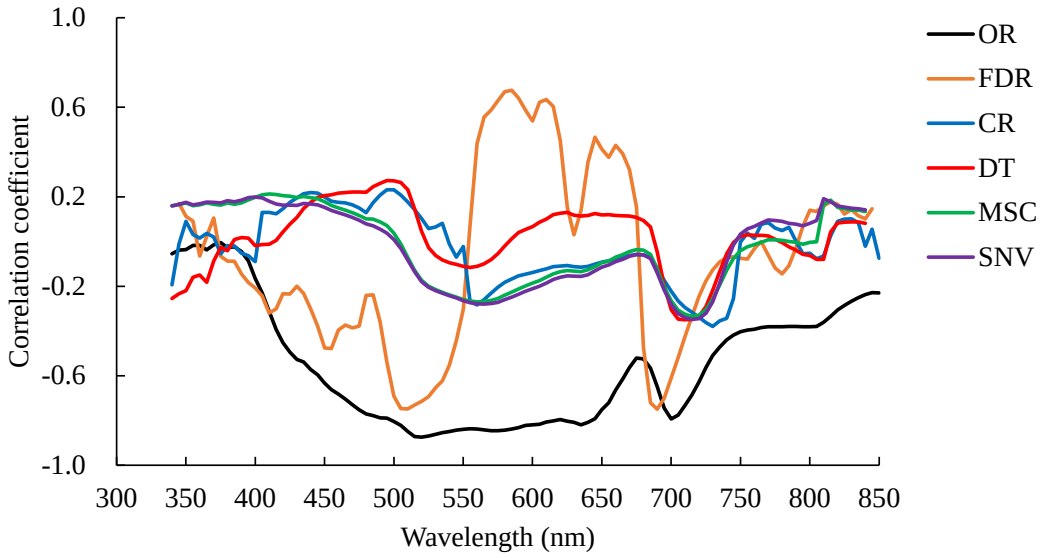

Figure 4. Data-based sensitivity analysis (DSA) results for six machine learning algorithms: Cubist, deep belief net (DBN), kernel-based extreme learning machine (KELM), random forests (RF), stochastic gradient boosting (SGB) and support vector machine (SVM). Importance values were calculated from 100 repetitions.

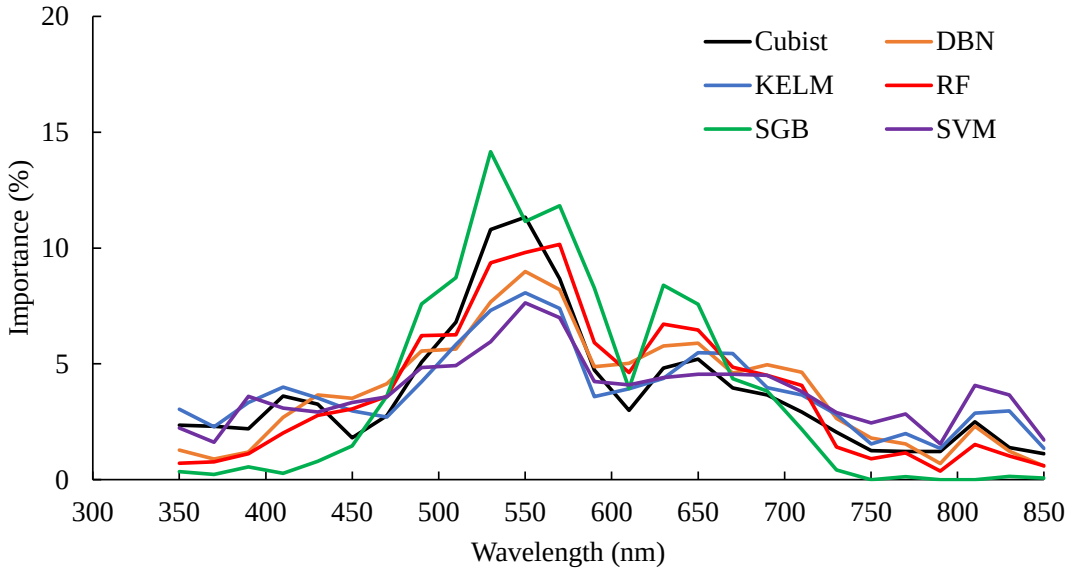

Table 1. Specification of the compact spectrometer, C12880MA-10

Parameter	Specification
Spectral response range	340 - 850 nm
Upper limit of calibrated full width at half maximum	15 nm
Bands	288
Dynamic range	16 bits
Dimension	20.1 ×12.5 ×10.1 mm
Weight	5 g

Table 2. Software and packages used for applying machine learning algorithms

Pre-processing/Machine learning algorithm	Software	Package
First derivative reflectance (FDR)	R	prospectr (Stevens and Ramirez-Lopez 2020)
Continuum removal (CR)	R	hsdar (Lehnert 2020)
De-tending (DT)	R	prospectr (Stevens and Ramirez-Lopez 2020)
Multiplicative Scatter Correction (MSC)	R	mdatools (Kucheryavskiy 2020)
Standard Normal Variate (SNV)	R	prospectr (Stevens and Ramirez-Lopez 2020)
Cubist	R	Cubist (Kuhn 2020)
Deep Belief Nets (DBN)	R	darch (Drees et al. 2015)
Kernel-based Extreme Learning Machine (KELM)	Matlab	Original code was downloaded from https://www.ntu.edu.sg/home/egbhuang/elm_codes.html.
Random Forests (RF)	R	randomForest (Breiman et al. 2018)
Stochastic Gradient Boosting (SGB)	R	gbm (Greenwell et al. 2020)
Support Vector Machine (SVM)	R	kernlab (Karatzoglou et al. 2019)

Table 3. RPD (ratio of performance to deviation) values of regression models from 100 repetitions. The pre-processing techniques are original reflectance (OR), first derivative reflectance (FDR), continuum-removal (CR) transformation, de-trending (DT), multiplicative scatter correction (MSC) and standard normal variate (SNV) and the machine learning algorithms are deep belief net (DBN), kernel-based extreme learning machine (KELM), random forests (RF), stochastic gradient boosting (SGB) and support vector machine (SVM)

	Cubist	DBN	KELM	RF	SGB	SVM
OR	2.01	1.80	1.82	1.97	1.78	1.38
FDR	1.40	1.30	1.40	1.50	1.38	1.11
CR	1.03	1.02	0.95	1.06	0.98	0.99
DT	1.04	1.02	0.89	1.06	1.00	0.88
MSC	0.99	1.02	0.91	1.03	0.92	0.99
SNV	1.03	1.02	0.94	1.06	0.99	0.97

Table 4. RMSE (root-mean-square error) values (μg cm-2) of regression models from 100 repetitions. The pre-processing techniques are original reflectance (OR), first derivative reflectance (FDR), continuum-removal (CR) transformation, de-trending (DT), multiplicative scatter correction (MSC) and standard normal variate (SNV) and the machine learning algorithms are deep belief net (DBN), kernel-based extreme learning machine (KELM), random forests (RF), stochastic gradient boosting (SGB) and support vector machine (SVM)

	Cubist	DBN	KELM	RF	SGB	SVM
OR	3.80	4.23	4.20	3.87	4.28	5.54
FDR	5.44	5.86	5.43	5.08	5.54	6.85
CR	7.42	7.46	8.03	7.21	7.77	7.69
DT	7.37	7.45	8.62	7.22	7.65	8.65
MSC	7.71	7.49	8.39	7.44	8.26	7.70
SNV	7.44	7.49	8.08	7.17	7.73	7.84