PLOS ONE

Predicting leaf nitrogen content in wolfberry trees by hyperspectral transformation and machine learning for precision agriculture

Yongmei Li[1,2]**, Hao Wang**[1,3,4]*****, Hongli Zhao**[3]*****, Ligen Zhang**[5]

1 School of Civil and Hydraulic Engineering, NingXia university, Yinchuan, Ningxia, People's Republic of China, **2** Institute of Agricultural Economy and Information Technology, Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan, Ningxia, People's Republic of China, **3** Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing, People's Republic of China, **4** State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, Beijing, People's Republic of China, **5** Ningxia Academy of building Research Co., Ltd, Yinchuan, Ningxia, People's Republic of China

* wanghao@iwhr.com (HW); zhaohl@iwhr.com (HZ)

Abstract

Leaf nitrogen content (LNC) is an important indicator for scientific diagnosis of the nutrition status of crops. It plays an important role in the growth, yield and quality of wolfberry. This study aimed to develop new spectral indices (NSIs) and constructed machine learning regression (MLR) models for predicting wolfberry tree LNC. By utilizing four smoothing methods and five mathematic transformation methods, we obtained the original spectral dataset and five spectral transformation datasets for quantitative analysis and model establishment. Subsequently, published vegetation indices (PVIs) were acquired, sensitive wavelengths (SWs) were screened and NSIs were calculated based on SWs. Then MLR models were constructed by combining NSIs from six spectral datasets with three machine learning algorithms. Finally, a comparison was made among the MLR models. The study indicated that the application of mathematical transformation highlighted the differences in spectra, the square root, first-derivative and second-derivative transformation improved the prediction accuracy of MLR models constructed based on NSIs (MLR-NSIs models). However, these transformations had little impact on improving the prediction ability of MLR models constructed based on PVIs (MLR-PVIs models). Additionally, The optimal model for predicting the LNC of wolfberry tree was obtained by using the Random Forest (RF) algorithm combined with NSIs developed by first-derivative transformation spectra. The determination coefficient of validation (R_v^2) and ratio of percentage deviation (RPD) was 0.71 and 1.90, respectively. In conclusion, this study has demonstrated that the combination of hyperspectral transformation and machine learning is useful for improving the accuracy of LNC estimation in wolfberry trees.

Introduction

Wolfberry is an important economic crop with the most local characteristics and advantages, and it becomes a pillar industry for local ecological governance and poverty alleviation in

Ningxia Province, China. In 2021, the area of wolfberry trees accounted for about 28666.7 ha in Ningxia Province, with a production of 25 billion RMB. By 2027, Ningxia plans to double the cultivated area and output value of wolfberry. Therefore, the high-quality development of wolfberry in Ningxia province receives much attention.

Nitrogen is an essential element for plant growth [1,2] and plays an important role in crop yield and quality [3]. Also, it is the mineral nutrient most commonly used in wolfberry orchard fertilization. Due to the high economic benefits, farmers usually use excessive chemical fertilizers in the planting of wolfberry. However, excessive nitrogen not only increases agricultural costs and pollutes the environment and underground water [4,5] but also affects crop yield and quality [6,7]. Currently, precision water and fertilizer integration technologies are implemented to reduce economic costs,mitigate environmental impacts, and promote the high-quality and sustainable development of the wolfberry industry. Evaluating wolfberry tree LNC is important and necessary to provide a scientific reference for better nitrogen nutrition administration by using water and fertilizer integration technologies.

Traditional methods commonly used for diagnosing crop nitrogen status include appearance diagnostic and chemical diagnostic methods. The appearance diagnostic methods involve assessing plant leaf color, growth patterns and symptoms. However, these method are subjective and prone to inaccuracy due to its reliance on visual identification of characteristic symptoms by individuals with varying levels of experience [8]. Some chemical diagnostic methods of laboratory determination based on field sampling are widely used to analyze the nitrogen or photosynthetic pigment content of plants, and they are proven to be accurate and repeatable, but these methods are destructive, time-consuming and laborious [9,10]. Additionally, traditional nitrogen estimates provide limited information due to sampling being restricted to a small number of sites within a given field [11,12]. Therefore, efficient alternatives are necessary. Hyperspectral remote sensing technology provides a new way to solve this problem owing to its high spectral resolution, simplicity, effectiveness, and non-destructiveness. Near-earth remote sensing offers a reliable method for data collection in precision agriculture through spectral monitoring [13].

There have been many studies on hyperspectral techniques for predicting nitrogen and chlorophyll levels in plants. such as wheat [14–17], rice [18,19], maize [20–22], apple trees [23] and olive trees [11]. Recently, the combination of hyperspectral feature parameters derived from transformation spectra and machine learning algorithms is being utilized to develop predictive models with high accuracy [24,25]. Previous studies have shown the feasibility of using hyperspectral remote sensing for monitoring plant nutrition, and hyperspectral remote sensing technology has become a major development trend in monitoring the nitrogen content of crops[26]. Despite the wide literature on nitrogen estimation by spectral measurements, very few studies are related to this work.

LNC is the most important indicator for evaluating the growth status of wolfberry trees. The timely and effective observation of nitrogen level is essential for guiding fertilization and controlling the growth, yield and quality of wolfberry [27,28]. However, to the best of our knowledge, as for the wolfberry tree, there is no study on predicting wolfberry LNC by hyperspectral data. In this study, our aims are to develop some NSIs and to construct the models for predicting wolfberry tree LNC by combining transformation hyperspectral data with machine learning algorithms. To achieve this goal, the following sub-objectives need to be addressed:

Testing four different smoothing methods (Savitzky–Golay smoothing, fast Fourier transform, continuous wavelet transform and nine-point weighted moving average method) to reduce the noise of the reflectance.

Using five mathematical transformations (reciprocal, logarithm, square root, first-derivative and second-derivative) to obtain spectral transformation datasets and analyzing the potential of the spectral datasets for estimating the LNC in Wolfberry trees.

Applying a wavelength combination-based method (difference, ratio and normalization) to develop new spectral indices (NSIs) that improve the fit of published vegetation indices (PVIs) to the LNC in wolfberry trees.

Comparing the predictive power for wolfberry tree LNC of three machine learning regression (MLR) models constructed based on PVIs and NSIs.

Materials and methods

Study area

The fieldwork was conducted from June to August 2019 in a wolfberry orchard located in Xinbu town, Zhongning County (105°26'~106°7'E, 37°9'~37°50'N), Ningxia Province, China. Its climate is continental arid and semi-arid in the north temperate zone, with an average annual temperature of 11.3°C and mean annual precipitation of 213 mm. The frost-free season is from 200 d to 205 d, the total annual sunshine is about 3476 h, and the average wind speed is 2 m/second. The terrain is hilly mixed with some alluvial valley plains, with an average altitude of 1140~1600 meters.

The study site has an area of 6.67 ha. According to the Chinese Soil Taxonomy System (1992), the soil type is mainly sandy loam soil. A drip irrigation system is used in this study site with water and nitrogen fertilizer integra-tion. All the wolfberry trees are planted with a fixed spacing in the spring of 2015. The row spacing and the plant spacing of the wolfberry trees are 3 m and 1m, respectively.

Sample collection

Ningqi No.7, one of the most widely cultivated varieties in Ningxia Province, was selected as the research object in this study. 52 healthy wolfberry trees were chosen as samples from the study area. To determine the nutritional status of wolfberry trees, a leaf analysis was conducted. Due to the variations in branch growth direction of wolfberry trees, leaves were randomly collected from new shoots in the east, south, west and north of each tree to represent the overall nutritional status. Each sample consisted of 100 g of healthy leaves collected from four orientations, with approximately 25 g of leaves obtained from each direction. A total of 52 leaf samples were gathered. Then, the leaves were placed in bags. The bags were sealed, labeled, placed in small mobile refrigerators filled with ice packs and brought back to the laboratory.

Leaf spectral reflectance measurement

The spectral reflectance was measured using an ASD FieldSpec 3 portable field spectrometer (Analytical Spectral Devices Inc., Boulder, USA) in a darkroom. The leaves arranged in a pile on a black mat with a thickness of 1 cm. A 50-Watt halogen lamp served as the light source, entering at a constant angle (45° from horizontal) and position-ed 30 cm away from the center of the leaf sample. The probe, with a field angle of 25°, was placed perpendicular to the surface of the leaf sample at a distance of 10 cm. The spectral reflectance for every leaf sample was determined as the mean of 10 sample lines. Additionally, a white reference panel was utilized for reflectance correction to ac-count for fluctuations in the light source.

LNC measurement

After measuring the spectra, all leaf samples were immediately placed in an oven at 104°C for 30 min, dried to a constant weight at 70°C, then ground into powder and passed through a

Table 1. The nitrogen content of the leaf samples. The "Total" represents the collection of all wolfberry leaf samples (n = 52) in this study. The "Calibration set" and "Validation set" represent wolfberry samples used for model construction and model validation respectively.

Sample	n	Maximum	Minimum	Mean	Median	standard deviation	variance	Coefficient of variation
		g kg^{-1}						
Total	52	4.46	3.53	4.098	4.155	0.224	0.050	0.055
Calibration set	35	4.34	3.53	4.090	4.16	0.222	0.049	0.054
Validation set	17	4.46	3.77	4.114	4.11	0.236	0.056	0.057

https://doi.org/10.1371/journal.pone.0306851.t001

0.25-mm screen. Subsequently, 0.1 g of powder extracted from each sample were used to determine nitrogen content. The nitrogen content was determined using a Kjeltec™ 8400 nitrogen analyser (FOSS, Hillerød, Denmark). In total, 52 nitrogen data were obtained. The foliar analyses (n = 52) indicated that LNC varied between 3.53% and 4.46%, with an average concentration of 4.098±0.224. The total samples are divided into a calibration set of 35 samples and a verification set of 17 samples randomly. The statistical results are shown in Table 1.

Spectral data pre-processing

In the signal acquisition process, spectra are often prone to different interferences due to the external environ-ment, the instrument's error, the structural and physical properties of the sample [11,29]. Therefore, to avoid the undesired influence on measured spectra, noise reduction is indispensable [30]. The common method to eliminate data noise is spectral smoothness. In this study, the raw data were smoothed by using the following methods: Savitzky–Golay Smoothing(S-G) [31], Fast Fourier Transform (FFT) [32], Continuous Wavelet Transform (CWT) [33] and Nine-point Weighted Moving Average (NWMA) [34]. The smoothed spectral data were labeled as the original spectra (R).

It was impossible to quantitatively judge the quality of smoothing methods by visual discrimination, so the spectral smoothness index (SSI) [35] was adopted in this study to evaluate the smoothness effect of spectral curves. The calculation of SSI is shown in Formula (1):

$$SSI = \frac{\sum \left(\lambda'_{i+1} - \lambda'_i\right)^2}{\sum \left(\lambda_{i+1} - \lambda_i\right)^2} \tag{1}$$

where SSI represents the spectral smoothness index; λ_i and λ'_i represent the spectrum of wavelength i before spectral smoothness and after spectral smoothness, respectively. The smaller the SSI, the smoother the whole spectral curve.

Mathematical transformation is one of the most effective analytic techniques for hyperspectral data [36]. It can highlight spectral features [37] and extract vegetation biochemical information [38,39]. In this study, five types of mathematical transformations including reciprocal (1/R), logarithmic(logR), square root(R$^{1/2}$), the first-derivative(R′) and the second-derivative (R″) transformations were performed on the original spectra and the corresponding spectral datasets were obtained. In the follow-up study, the original spectral dataset and five spectral transformation datasets were used as the basis datasets for quantitative analysis and model establishment.

Hyperspectral features

To explore the spectral indices applicable for LNC prediction of wolfberry tree, we gathered as many published vegetation indices (PVIs) as possible, while also developing new spectral indices (NSIs). Table 2 summarizes the vegetation indices used by other studies to determine the nitrogen status of crops.

Table 2. Published vegetation indices used by other studies.

Vegetation Indices	Calculation formula	Reference
RVI	R_{800}/R_{670}	[40]
NDVI	$(R_{800}-R_{670})/(R_{800}+R_{670})$	[41]
VOG1	R_{740}/R_{720}	[42]
VOG3	$(R_{734}-R_{747})/(R_{715}+R_{720})$	[43]
MTCI	$(R_{750}-R_{710})/(R_{710}-R_{680})$	[44]
GM	R_{750}/R_{700}	[45]
SR705	R_{750}/R_{705}	[46]
mSR705	$(R_{750}-R_{445})/(R_{705}-R_{445})$	[47]
ND705	$(R_{750}-R_{705})/(R_{750}+R_{705})$	[46]
mND705	$(R_{750}-R_{705})/(R_{750}+R_{705}-2R_{445})$	[47]
$CI_{red\ edge}$	$R_{750}/R_{720}-1$	[48]
$NDVI_{gb}$	$(R_{573}-R_{440})/(R_{573}+R_{440})$	[49]
VI_{opt}	$(1+0.45)(R_{800}^2+1)/(R_{670}+0.45)$	[50]
NDRE	$(R_{790}-R_{720})/(R_{790}+R_{720})$	[51]
MCARI	$(R_{700}-R_{670}-0.2(R_{700}-R_{550}))(R_{700}/R_{670})$	[52]
MTVI2	$1.5(1.2(R_{800}-R_{550})-2.5(R_{670}-R_{550}))/sqrt((2R_{800}+1)^2-(6R_{800}-5sqrt(R_{670}))-0.5)$	[53]
CVI1	MCARI/MTVI2	[54]
PPR	$(R_{550}-R_{450})/(R_{550}+R_{450})$	[55]
PRI	$(R_{570}-R_{531})/(R_{570}+R_{531})$	[56]
GNDVI	$(R_{750}-R_{550})/(R_{750}+R_{550})$	[57]
NPCI	$(R_{430}-R_{680})/(R_{430}+R_{680})$	[58]
NRI	$(R_{570}-R_{670})/(R_{570}+R_{670})$	[59]
SIPI	$(R_{810}-R_{460})/(R_{810}+R_{460})$	[60]
TVI	$0.5[120(R_{750}-R_{550})-200(R_{670}-R_{550})]$	[61]
VARI	$(R_{550}-R_{670})/(R_{550}+R_{670}-R_{445})$	[62]
TGI	$-0.5[(\lambda_{670}-\lambda_{480})(R_{670}-R_{550})-(\lambda_{670}-\lambda_{550}))(R_{670}-R_{480})]$	[63]
TCARI	$3((R_{700}-R_{670})-0.2(R_{700}-R_{550})(R_{700}/R_{670}))$	[64]
OSAVI	$1.16(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)$	[65]
CVI2	TCARI/OSAVI	[64]
DCNI	$(R_{720}-R_{700})/(R_{700}-R_{670})/(R_{720}-R_{670}+0.03)$	[15]
$SR_{550,670}$	R_{550}/R_{670}	[66]
$SR_{780,550}$	R_{780}/R_{550}	[66]
$SR_{780,670}$	R_{780}/R_{670}	[66]
$NDI_{780,670}$	$(R_{780}-R_{670})/(R_{780}+R_{670})$	[66]
REP	$(R_{670}+R_{780})/2$	[67]

Mltivariable stepwise regression (MSR) is a simple and effective method to avoid multicollinearity among factors and screen characteristic variables [68]. Therefore, it was employed to screen sensitive wavelengths (SWs) that can represent the LNC of wolfberry tree. The screening SWs was performed through forward selection by using the Scientific Platform Serving for Statistics Professional (SPSSPRO) in china. The reflectance values of the original, reciprocal, logarithmic, square root, first-derivative, and second-derivative spectra within the range of 350 nm to 2500 nm were considered as independent variables, with LNC being the dependent variable. The significance level of the selected variable was set to 0.01 and the excluded variable was set to 0.05. Subsequently, new spectral indices (NSIs) were calculated based on the screened SWs. These NSIs included difference spectral index (DSI), ratio spectral index (RSI)

and normalized difference spectral index (NDSI). The NSIs are calculated as follows:

$$DSI = R_{\lambda i} - R_{\lambda j} \qquad (2)$$

$$RSI = R_{\lambda i}/R_{\lambda j} \qquad (3)$$

$$NDSI = (R_{\lambda i} - R_{\lambda j})/(R_{\lambda i} + R_{\lambda j}) \qquad (4)$$

where R represents the spectral reflectance, and λi represents the wavelength.

Machine learning regression model

Construction of the machine learning regression model. In this study, the hyperspectral feature was used as the independent variable, and three machine learning algori-thms, namely, Adaptive Boosting (AdaBoost), Extra Trees(ET) and Random Forest (RF) were used to construct LNC prediction models.

AdaBoost is an iterative algorithm that combines multiple weak classifiers (typically single decision stumps) to form a strong classifier [69]. Weak classifiers usually refere to classifiers with slightly better generalization ability than random guessing, while strong classifiers have a generalization ability very close to the true value. In AdaBoost, each round of iteration adjusts sample weights based on the previous learning results so that the misclassified samples receive more attention in the subsequent round. Simultaneously, the algorithm assign a weight to each weak classifier based on its performance on the current training set (giving high weight to classifiers with low rates and low weight to classifiers with high rates) [69,70]. All weak classifiers are combined in a weighted manner to form the final strong classifier. RF is an ensemble learning algorithm that consists of multiple decision trees. It is often used in classification and regression problems. And it produces good results for both linear and nonlinear data [71]. ET is a relatively new machine-learning algorithm and is developed as an extension of the random forest algorithm. It employs the same principle as RF and uses a random subset of features to train each base estimator [72]. Meanwhile, it uses the whole training dataset to train each regression tree. In contrast, RF uses a bootstrap replica to train the model [72].

Robustness and model performance analysis. The application of multiple evaluation indicators can provide insights into the overall accuracy, variability, and explanatory power of the model. In this study, the determination coefficient (R^2), the root mean square error (RMSE), mean absolute error (MAE) and ratio of percentage deviation (RPD) were used to evaluate the robustness of the prediction models. R^2 is a statistical measure of the correlation between variables, with values ranging from 0 to 1[73]. The larger R^2 was, the better the fit between the predicted values and the observed values. RMSE is the square root of Mean Square Error (MSE) which is the average of the squared differences between the predicted values and the observed values [73]. MAE measures the average absolute difference between predicted values and observed values [74]. The smaller RMSE and MAE are indicating that predicted values are closer to observed value. RPD is defined as a ratio of the standard deviation (SD) of the laboratory measured data to the root-mean-square error (RMSE) of prediction [75,76]. Therefore, when R^2 approaches 1 while RMSE and MAE approach 0, it indicates better predictive ability for a model. An RPD value less than 1.4 suggests poor prediction quality; an RPD value between 1.4 and 1.8 indicates fair prediction suitable for assessment; an RPD value between 1.8 and 2.0 suggests good prediction where quantitative predictions are possible; an RPD value larger than 2.0 indicates very good ability to estimate samples[77]. The robustness of the prediction models was evaluated by:

determination coefficient of calibration (R_c^2);

determination coefficient of validation (R_v^2);
root mean square error of calibration ($RMSE_c$);
root mean square error of validation ($RMSE_v$);
mean absolute error of calibration (MAE_c);
mean absolute error of validation (MAE_v);
ratio of percentage deviation (RPD).
The R^2, RMSE, MAE and RPD are given as follows:

$$R^2 = \left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \right]^2 \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{n}} \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \tag{7}$$

$$RPD = \frac{SD}{RMSE} \tag{8}$$

where X_i and Y_i represent the estimated and measured values, respectively. n represents the number of samples. $\bar{X}$ and $\bar{Y}$ represent the average of estimated and measured values, respectively. SD and RMSE are the standard deviation and root mean square error of the validation set, respectively.

Results

Spectral data preprocessing

The raw spectra of 52 wolfberry leaf samples are smoothed by using S-G, FFT, CWT and NWMA methods and the mean reflectance of four smoothed spectra is calculated, respectively. The four smoothing methods have similar effects on spectral curves. It is difficult to determine which smoothing method worked best by visual discrimination. Therefore, SSI is used to evaluate the smoothing quality. FFT has the smallest SSI, so the spectra smoothed by FFT are selected as the original spectra(R) in this study (S1 Fig).

Six types of spectral curves (S1 Fig) are shown in Fig 1. Compared with the original spectra, reciprocal, logarithmic and square root transformations all enhance the difference of spectral reflectance in the VIS and in the SWIR region to varying degrees. Meanwhile, the first-derivative and second-derivative transformation effectively improve the resolution of the overlapping spectra, so the difference at the red-edge position (680~760nm) is increased. Therefore, the mathematical transformation can improve the spectral features to a certain extent, which helps to distinguish the impact of the difference in nitrogen content on the spectra.

Hyperspectral features analysis

Relationship of PVIs to LNC.   Fig 2 represents the correlation coefficient between PVIs and LNC(S2 Fig). In the original spectra, the 24 PVIs (accounting for 69%) pass the 0.01-significance test and are significantly correlated with LNC. The absolute values of the correlation coefficients of PRI, REP, and VOG3 are the largest, which are -0.544,-0.528 and -0.521,
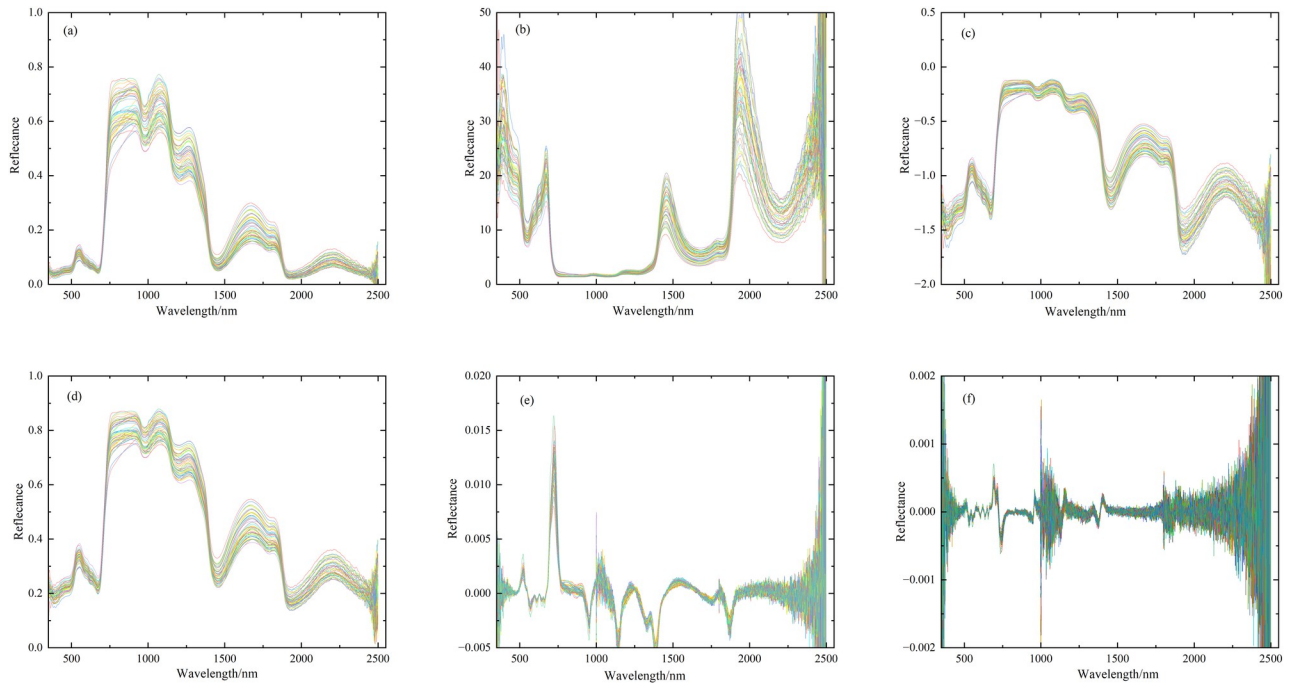
Fig 1. The curves of the original spectra and five types of transformation spectra. Each color line represents a sample reflectance curve, with a total of 52 samples. (a) represents the original spectra (R); (b), (c), (d), (e) and (f) represent the reciprocal (1/R), logarithmic (logR), square root ($R^{1/2}$), first-derivative (R') and second-derivative spectra (R''), respectively.

https://doi.org/10.1371/journal.pone.0306851.g001

respectively. In reciprocal spectra, there are19 PVIs with a significance level of 0.01, accounting for 54% of all PVIs. The absolute values of the correlation coefficients of PRI, VOG3 and GM are the largest, which are -0.535, -0.511 and 0.487, respectively. In the logarithm spectra, 23 PVIs pass the 0.01-significance test, accounting for 66%. The absolute value of the correlation coefficient of VOG3 is the largest, which is -0.525. Followed by PRI and GM, with the correlation coefficients of -0.519 and 0.507, respectively. In the square root spectra, 26 PVIs pass the 0.01-significance test, accounting for 74%. PRI has the largest correlation coefficient of 0.535. The mSR_{705} and mND_{705} have the second highest correlation coefficients, both with value of -0.515. In the first-derivative and second-derivative spectra, 3 and 9 PVIs pass the 0.01-significance test, accounting for 37% and 26%, respecttively. In the first-derivative spectra, the top three PVIs with the absolute value of the correlation coefficients are GM, TVI and $CI_{red\ edge}$, with values of -0.497, -0.489 and -0.442. In second-derivative spectra, the top three PVIs are
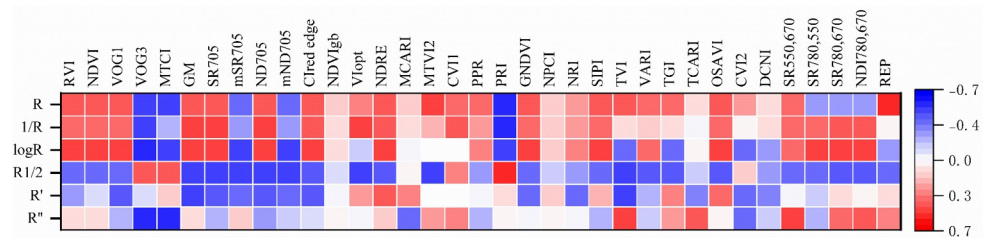

Fig 2. The correlation coefficient between PVIs with LNC in different spectral datasets. R, 1/R, logR, $R^{1/2}$, R' and R'' represent the original spectra, reciprocal spectra, logarithmic spectra, square-root, first-derivative and second-derivative spectra, respectively.

https://doi.org/10.1371/journal.pone.0306851.g002

Table 3. The sensitive wavelengths selected by MSR. R, 1/R, logR, $R^{1/2}$, R' and R'' represent the original spectra, reciprocal spectra, logarithmic spectra, square-root, first-derivative and second-derivative spectra, respectively. The selected sensitive wave-lengths rank according to wavelength.

spectra	Sensitive wavelengths (nm)
R	359, 772
1/R	359, 771
logR	359, 770
$R^{1/2}$	359, 771
R'	379, 388, 449, 745, 1128, 1610, 1945, 2005, 2051, 2073, 2092, 2102, 2109, 2153, 2240, 2442
R''	759, 774, 787, 1009, 1028, 1147, 1454, 1518, 1585, 1641, 1643, 1796, 1950, 2202, 2359

https://doi.org/10.1371/journal.pone.0306851.t003

VOG3, MTCI and TVI, with values of -0.561, -0.538 and 0.502. Overall, the spectral transformation has little effect on improving the correlation between PVIs and LNC. Only there is the VOG3 in second-derivative spectra whose correlation coefficient is slightly greater than that of the original spectra.

Relationship of NSIs to LNC. *Sensive wavelengths selection.* The selected SWs are presented in Table 3. The SWs of the original, reciprocal, logarithmic and square root reflectance are observed at 359 nm in the VIS region and near 770nm in the NIR region. A total of 16 SWs are identified from the first-derivative spectra, including 379nm, 388nm and 449nm distributed in the VIS region, as well as 745 nm and 1128nm distributed in the NIR region and the remaining SWs distributed in the SWIR region. However, the SWs of the second-derivative reflectance don't appear in VIS region but are instead distributed in the NIR and SWIR region. Upon comparing the SWs from different spectral datasets, it is found that the first-derivative transform spectra have the higher number of SWs and cover the wider spectral range.

Relationship of NSIs to LNC. NSI was calculated based on the SW_S data in Table 3 resulting in a total of 687 NSIs. Correlation analysis between NSIs and LNC was conducted. As indicated in Table 4, the original spectra, the logarithmic spectra and the square-root spectra all show that all NSIs pass the 0.01-significance test. The NSIs with the strongest correlation are $NDSI_{R(772,359)}$, $DSI_{logR(770,359)}$ and $DSI_{R^{1/2}(771,359)}$, with a correlation coefficient of -0.556, -0.563 and -0.599, respectively. In the reciprocal spectra, $RSI_{1/R(771,359)}$ and $NDSI_{1/R(771,359)}$ pass the 0.01-significance test while $DSI_{1/R(771,359)}$ only passes the 0.05-significance test. In the first-derivative spectra, a total of 51 NSIs are correlated with LNC at a significance level of 0.01 and the top three NSIs are $DSIR'_{(1128,388)}$, $DSIR'_{(745,449)}$ and $DSIR'_{(1945,745)}$, with correlation coefficients of 0.597, -0.582 and 0.573, respectively. In the second-derivative spectra, a total of 29 NVIs pass the 0.01-significance test and the top three NSIs are $DSIR''_{(1641,787)}$, $DSIR''_{(1641,759)}$ and $DSIR''_{(1518,759)}$, with the correlation coefficients of -0.644, -0.600 and -0.551, respectively.

The predicting model

Machine learning regression models based on PVIs. To assess the feasibility of using PVIs to monitor wolfberry tree LNC and to evaluate the impact of spectral transformations on the response of PVIs to wolfberry tree LNC change. The top three PVIs significantly correlated with LNC were selected as independent variables to establish machine learning regression models for predicting the LNC of wolfberry trees (S3 Fig). Fig 3 illustrates that among the three types of MLR-PVIs models, the Ada-Boost models exhibites the highest R_c^2, with a value exceeding 0.987. The R_c^2 values of RF and ET models are also high and similar, ranging from 0.837 to 0.891 and from 0.841 to 0.892, respectively. When considering six types of spectral

Table 4. The correlation coefficients of all new spectral indices in the original spectra, the reciprocal spectra, logarithmic spectra and square root spectra. And the top three NSIs significantly correlated with LNC in first-derivative spectra and second-derivative spectra. The correlation coefficient refers to the result of correlation analysis between LNC and NSIs calculated by SWs from different spectral datasets. DSI, RSI and NDSI represent the difference spectra index, ratio spectral index and normalized difference spectral index respectively. R, 1/R, logR, $R^{1/2}$, R' and R'' represent the original spectra, the reciprocal spectra, logarithmic spectra, square root spectra, first-derivative spectra and second-derivative spectra respectively. The NSIs are listed in descending order of the absolute values of the correlation coefficients.

spectra	NSIs	Correlation Coefficients	spectra	NSIs	Correlation Coefficients
R	$NDSI_{R(772,359)}$	-0.556**	$R^{1/2}$	$DSI_{R^{1/2}(771,359)}$	-0.599**
	$RSI_{R(772,359)}$	0.554**		$NDSI_{R^{1/2}(771,359)}$	-0.562**
	$DSI_{R(772,359)}$	-0.467**		$RSI_{R^{1/2}(771,359)}$	0.560**
1/R	$NDSI_{1/R(771,359)}$	0.556**	R'	$DSIR'_{(1128,388)}$	0.597**
	$RSI_{1/R(771,359)}$	-0.554**		$DSIR'_{(745,449)}$	-0.582**
	$DSI_{1/R(771,359)}$	0.334*		$DSIR'_{(1945,745)}$	0.573**
logR	$DSI_{logR(770,359)}$	-0.563**	R''	$DSIR''_{(1641,787)}$	-0.644**
	$NDSI_{logR(770,359)}$	0.501**		$DSIR''_{(1641,759)}$	-0.600**
	$RSI_{logR(770,359)}$	-0.410**		$DSIR''_{(1518,759)}$	-0.551**

** and * indicate a significant correlation at the levels of 0.01 and 0.05 respectively.

datasets, it is observed that the R_c^2 values for AdaBoost models constructed by using five transformation spectra closely resemble those constructed by using original spectra. However, Except for the first-derivative spectra, the R_c^2 values of the ET models constructed by using other transformation spectra are lower than those constructed by using the original spectra. Furthermore, it is only in the case of the first-derivative or logarithmic spectra were utilized that the R_c^2 value of the RF models slightly exceed those obtained when utilizing original spectra. The above analysis shows that spectral transformation can hardly improve the prediction ability of the MLR-PVIs model.

As shown in Fig 4, the ET models of the three types of MLR-PVIs models exhibit the strongest explanatory ability for LNC with the R_v^2 values ranging from 0.23 to 0.58, followed by the RF models with the R_v^2 values ranging from 0.16 to 0.52 and then the AdaBoost models with the R_v^2 values ranging from 0.03 to 0.35. Among the six types of spectral datasets, both original spectra and logarithmic spectra utilized in constructing MLR-PVIs models show good explanatory ability for LNC change, with the R_v^2 values ranging from 0.35 to 0.58 and from 0.41 to
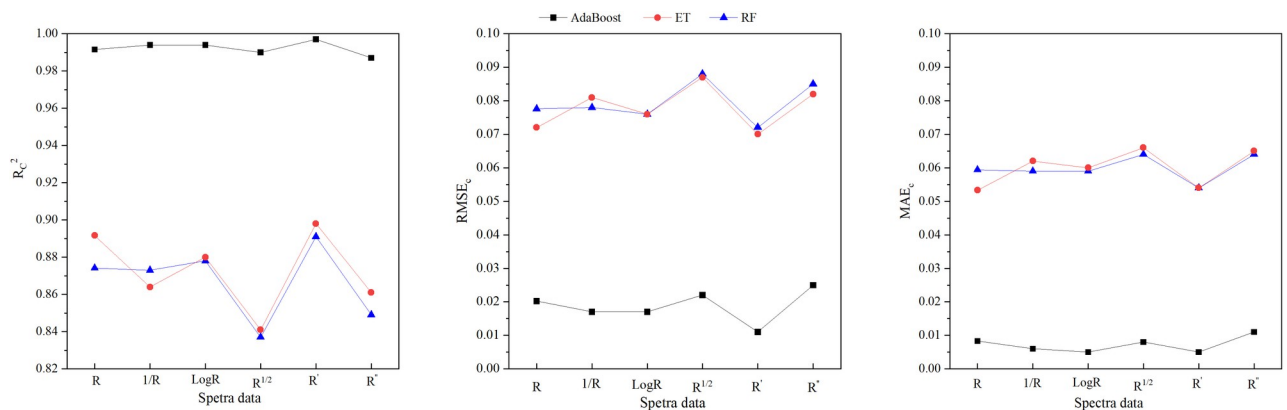

Fig 3. R_c^2, $RMSE_c$ and MAE_c distributions of the machine learning regression models based on PVIs (MLR-PVIs models) in differrent spectral dataset. R_c2, $RMSE_c$ and $MAEc$ represente determination coefficient, root mean square error and mean absolute error of calibration respecttively. R, 1/R, logR, $R^{1/2}$, R' and R'' represent original spectra, reciprocal spectra, logarithmic spectra, square root spectra, first-drivative and second-derivative spectra respectively. AdaBoost, ET and RF represent Adaptive Boosting models, Extra Trees models and Random forest models respectively.
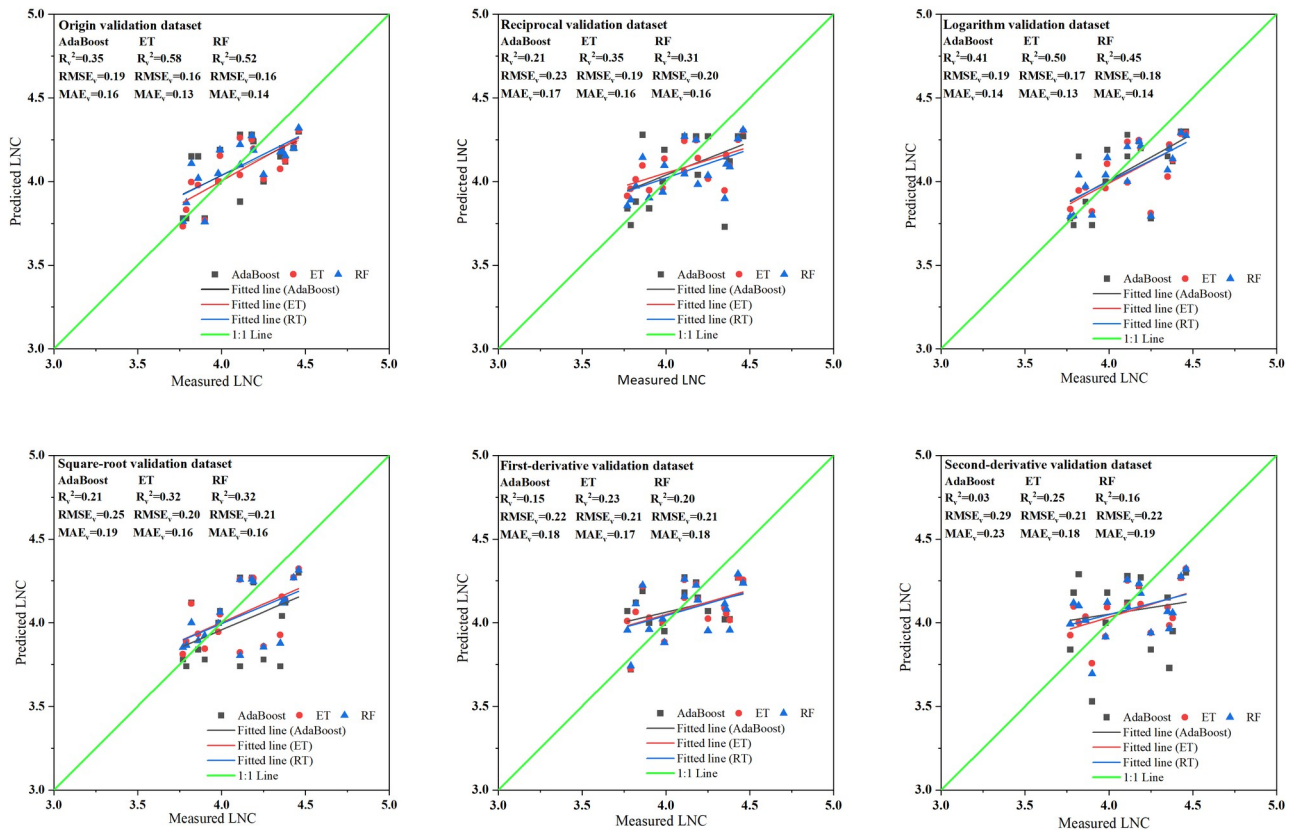
Fig 4. Distribution of measured and predicted values of the machine learning regression models based on PVIs (MLR-PVIs models) in different spectral datasets. AdaBoost, ET and RF represent Adaptive Boosting models, Extra Trees models and Random forest models respectively. R_v^2, $RMSE_v$ and MAE_v represent determination coefficient, root mean square error and mean absolute error of validation respectively.

0.50 respectively. However for the other four transformation spectra used in MLR-PVIs models, all show very poor predictive ability for wolfberry tree LNC, and the R_v^2 values ranging from 0.03 to 0.35 (S4 Fig). Overall, In the MLR-PVIs models, the ET model constructed by using the original spectra displays a good explanatory ability for LNC change, with the R_v^2 of 0.58, the $RMSE_c$ of 0.16 and the MAEc of 0.13. Followed by the RF model constructed by using the original spectra and the ET model constructed by using the logarithmic spectra and their R_v^2 values are 0.52 and 0.50, respectively. The analysis indicates that spectral transformations had minimal impact on improving the predictive power within MLR-PVIs models.

Machine learning regression models based on NSIs

The NSIs listed in the Table 4 were used to construct AdBoost, ET and RF models for predicting the wolfberry trees LNC. Fig 5 illustrates that among the three types of MLR-NSIs models, the AdaBoost models exhibit the highest R_c^2, with a value exceeding 0.985. The R_c^2 values of the RF and ET models are comparable, ranging from 0.868 to 0.922 and from 0.874 to 0.922, respectively (S5 Fig). When considering six types of spectral datasets, it is observed that R_c^2 values for AdaBoost models constructed by using five transformation spectra closely resemble those constructed by using original spectra. The R_c^2 values of ET and RF models constructed by using the five transformation spectra are greater than those constructed by using the original spectra.
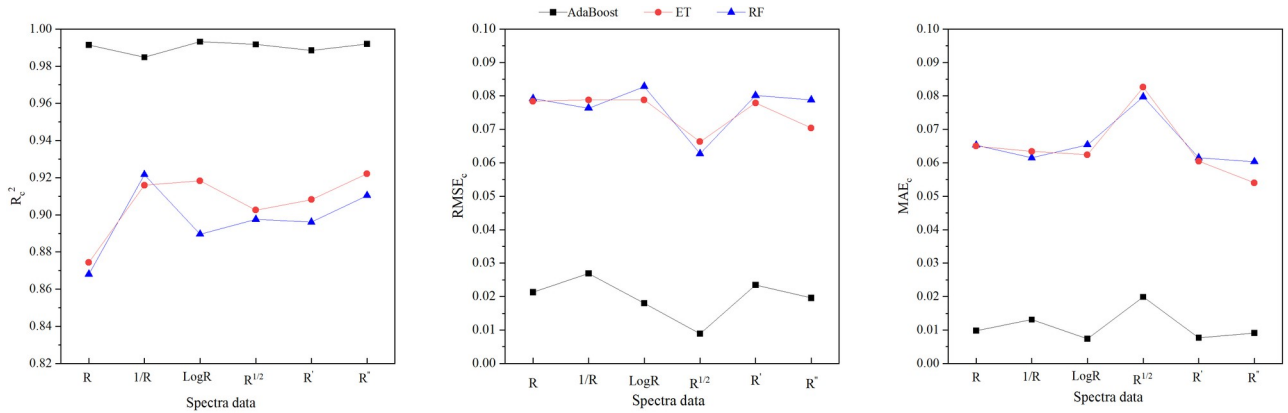
Fig 5. Rc2, RMSEc and MAEc distributions of the machine learning regression models based on NSIs in different spectral dataset. R_c^2, $RMSE_c$ and MAE_c represent determination coefficient, root mean square error and mean absolute error of calibration respectively. R, 1/R, logR, $R^{1/2}$, R' and R'' represent original spectra, reciprocal spectra, logarithmic spectra, square root spectra, first-derivative and second-derivative spectra respectively. AdaBoost, ET and RF represent Adaptive Boosting models, Extra Trees models and Random forest models respectively.

https://doi.org/10.1371/journal.pone.0306851.g005

As illustrated in Fig 6 (S6 Fig), the RF models of the three types of MLR-SIs models obtain the highest R_v^2, ranging from 0.433 to 0.710, followed by the AdaBoost models with the R_v^2 values ranging from 0.408 to 0.688, and then the ET models with the R_v^2 values ranging from 0.376 to 0.632. Among the six types of spectra datasets, it is found that the first-derivative spectra utilized in constructing MLR-PVIs models showed the strongest explanatory ability for LNC change, with the R_v^2 values ranging from 0.632 to 0.710. Followed by the second-derivative spectra with R_v^2 values ranging from 0.517 to 0.628, and then the square root spectra with R_v^2 values ranging from 0.475 to 0.563. However, for the logarithmic and the reciprocal spectra, it is observed that their R_v^2 values are lower than those of the original spectra, ranging from 0.377 to 0.534 and ranging from 0.376 to 0.500, respectively. The original spectra used in MLR-NSIs models shows R_v^2 values between 0.404 and 0.540.

Overall, in terms of MLR-NSIs models, the RF model constructed by using first-derivative spectra demonstrates superior explanatory power with a R_v^2 value of 0.710, a $RMSE_v$ value of 0.124 and a MAE_v value of 0.104. Followed by the AdaBoost model constructed by using the first-derivative spectra, with a R_v^2 value of 0.688, a $RMSE_v$ value of 0.130 and a MAE_v value of 0.108. Then the ET model constructed using the first-derivative spectra, with an R_v^2 value of 0.632, an $RMSE_v$ value of 0.141 and a MAE_v value of 0.111. Additionally, the ET model constructed by using the second-derivative spectra exhibits good explanatory power as well, with a R_v^2 value of 0.628. The analysis above indicates that the NSIs developed through spectral transformation can enhance prediction capabilities within MLR-NSIs models.

Model accuracy comparison

As shown in Fig 7 (S7 Fig), the RPD values of the MLR-PVIs models constructed by using the five transformation spectra are smaller than those constructed by using the original spectra. Among the five transformation spectra, the logarithmic spectra yield the highest RPD values, with a maximum value of only 1.355. In all MLR-PVIs models, only the ET model and RF model constructed by using the original spectra achieve an RPD value slightly greater than 1.400, with PRD values of 1.513 and 1.483 respectively.

When utilizing the reciprocal transformation spectra, the RPD values of the MLR-NSIs models are less than 1.4. By using the original spectra, the logarithmic spectra and the square root spectra, only the RF models achieve RPD value of slightly larger than 1.4, with RPD values
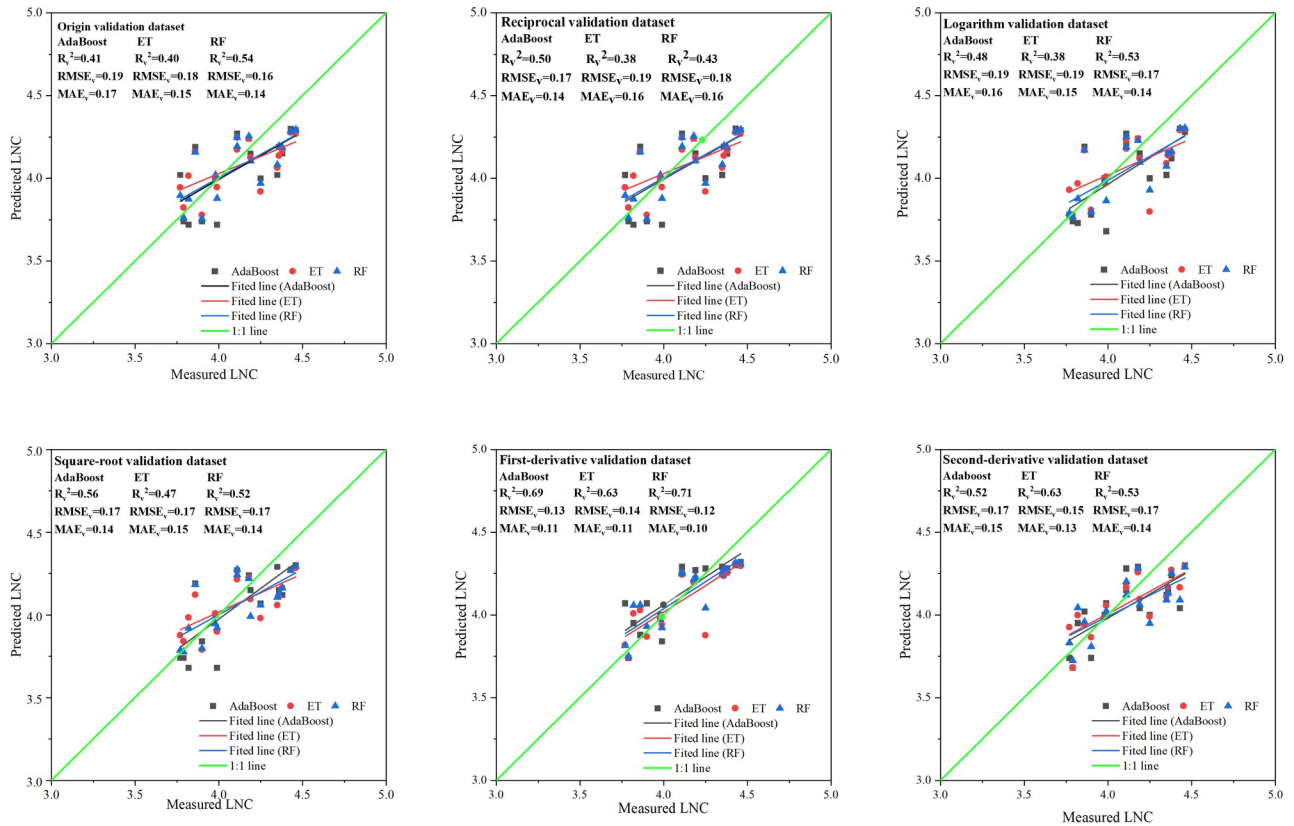
Fig 6. Distribution of measured and predicted values of the machine learning regression models based on NSIs (MLR-NSIs models) in di-fferent spectral datasets. AdaBoost, ET and RF represent Adaptive Boosting models, Extra Trees model and Random forest models respecttively. R_v^2, $RMSE_v$ and MAE_v represent determination coefficient, root mean square error and mean absolute error of valida-tion respectively.

of 1.456, 1.410 and 1.425 respectively. By using the second-derivative spectra, only the RPD value of the ET model was 1.589, all other models have PRD values below 1.400. By using the first-derivative spectra, the RPD values of the RF model, AdBoost model and ET model are1.902, 1.816 and 1.678 respectively. So the MLR-NSIs models constructed by using the first-derivative spectra are outstanding. Overall, the MLR-NSIs models constructed by using the transformation spectra demonstrate better predictive capabilities for wolfberry tree LNC compared to MLR-PVIs models-except for all models of the original spectra and the ET model of the logarithmic spectra.

Discussion

Hyperspectral data contains a large amount of redundant information due to the relatively few variables that really and effectively control the spectral signatures of vegetation [11,78]. Research has shown that while machine learning can effectively handle complex collinearity problems, the predicttive ability of models are influenced by the varying capabilities of different machine learning algorithms [79]. Additionally, studies have demonstrated that spectral mathematical transformation plays a crucial role in amplifying the differences in spectral reflectance and extracting spectral characteristics [80,81]. In order to identify suitable variables for wolfberry tree LNC and mitigate the impact of collinearity in hyperspectral data on the predictive ability of MLR models, mathematical transformations were applied to the original
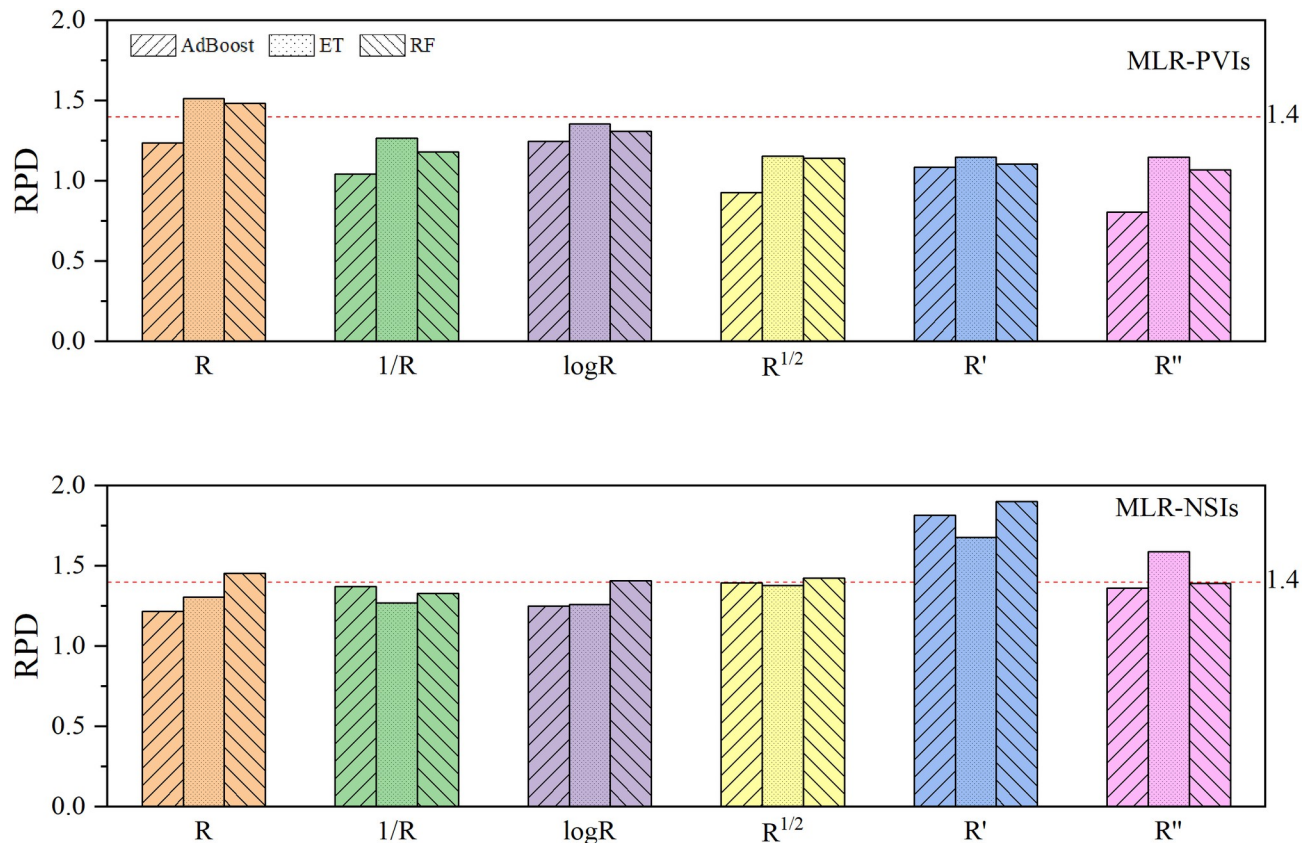
Fig 7. RPD distribution of the MLR-PVIs models and the MLR-NSIs models. The RPD represents the ratio of the standard deviation (SD) of the measured LNC to the root mean square error of predicted LNC. The MLR-PVIs models and the MLR-NSIs models represent the machine learning regression models constructed by using the published vegetation indices and by using the new spectral indices. AdaBoost, ET and RF represent Adaptive Boosting models, Extra Trees models and Random forest models respectively. R, 1/R, logR, $R^{1/2}$, R′ and R″ represent original spectra, reciprocal spectra, logarithmic spectra, square root spectra, first-derivative and second-derivative spectra respectively.

spectra. Subsequently, SWs were selected from six transformed spectral datasets and NSIs were calculated by combining with the SWs.

In this study, the reciprocal, logarithm and square-root transformation enhance the spectral differences of the VIS and SWIR region. The first-derivative and the-second derivative transformation highlight the subtle spectral differences in the red-edge region. The sensitive spectra located at 359nm, 379nm, 388nm and 449nm in the VIS region are selected.They might be directtly related to the chl a absorption pit at 430 nm and indirectly related to the N content [64,82]. The spectra located at 745nm, 759nm, 774nm, 787nm in the NIR region are selected. These spetra are close to 750nm and are in the limit between red domain of the VIS region and red-edge domain of the NIR region. Several authors selected the hyperspectral reflectance at 750 nm to estimate Chl ab content [83,84]. Therefore, it also could be related to N content. Meanwhile, the selected sensitive wavelengths are found to be consistent with previously published vegetation indices. For example, the sensitive wavelengths from first-derivative spectra (745nm and 449nm) closely match the wavelength of mSR705 (750nm and 450nm) in PVIs [47]. Studies have shown characteristic absorption bands of nitrogen are primarily located in the shortwave infrared (SWIR) region at specific wavelengths such as 1510nm, 1730nm, 1940nm, 1980nm, 2060nm, 2180nm, 2240nm, 2300nm and 2350nm [85]. However, the spectra within the SWIR region are sensitive to leaf water content and the indices which included

the 1770nm absorption feature are least affected by leaf water content [86]. The 1128nm, 1610nm, 1945nm, 2051nm, 2073nm, 2092nm, 2153nm and 2240nm selected from first-derivative spectra and the 1641nm, 1643nm, 1796nm, 1950nm, 2202nm, 2359nm selected from second-derivative spectra are located in the SWIR region, which might be related to N and water content [82]. The 1518nm and 1585nm selected from the second-derivative spectra fell within the spectral range of 1500-1600nm, which is a region dominated by absorption features caused by stretching of N-H bonds at 1510nm [82] and this bond is related to the amount of N present in protein [11]. Furthermore, research indicat that employing spectral mathematical transformation can enhance model accuracy by transforming linear relationships between spectral reflectance and observed objects into nonlinear ones [87]. In this study, spectral transformations have little impact on improving the correlation between PVIs and LNC and the predictive power of MLR-PVIs models. However, except for the ET models constructed base on the logarithmic transformation spectra, the response capacity of all the MLR-NSIs models constructed base on the transformation spectra to wolfberry tree LNC are improved compared with the MLR-PVIs models. And compared to the original spectra, the square root, first-derivative and second-derivative transformations have improved the predicttive ability of the MLR-NSIs models.

Many crop nitrogen monitoring models have utilized various types of multiple regressions techniques, including stepwise multiple linear regression [88,89], partial least squares regression [89], and machine learning regression (e.g., support vector machine regression, random forest regression, adaptive boost and bagging.et al) [79,89]. These nonlinear modeling methods can incorporate more spectral information to provide more accurate results. Hyper-spectral evaluation of nitrogen accumulation in winter wheat leaves based on different regression models shows that the model constructed by machine learning can effectively utilize the characteristic information of chlorophyll absorption and improve the estimation accuracy [26]. Ensemble machine learning models (RF, AdBoost and Bagging) have good predictive ability for crop nitrogen nutrition [79]. In this study, the three ensemble machine learning models (AdBoost, ET and RF) constructed based on first-derivative spectra all exhibit strong predictive performance for wolfberry tree LNC. Notably, the RF model constructed using first-derivative spectra demonstrates optimal predictive power. This outcome may also be attributed to the NSIs utilized in the model: $DSIR'_{(1128,388)}$, $DSI_{R'(745,449)}$ and $DSI_{R'(1945,745)}$. These three NSIs encompass spectral features at 449nm in the VIS region, 745nm in the red-edge region, 1128nm in the NIR region and 1945nm in the SWIR region. Spectral information significantly impact a model's predictive ability and an abundance of such information can enhance it to a certain extent.

Studies have shown that nitrogen content changes with the progression of growth stages [90,91]. Furthermore, it has been indicated that the small sample quantity is not a negligible factor that affects the robustness of the estimation model [26]. As our study was conducted using specific synthetic datasets composed of different growth stages and with limited data (only 52 samples), our models inherently possess limitations in applicability. Addi-tionally, previous studies have mainly focused on farmland with artificially controlled nitrogen levels where high and low N levels were easily obtained [91]. However, variable rate fertilization is rarely carried out in Chinese agriculture, making it challenging to obtain data with large changes in biochemical components as observed in the experiment [91]. Therefore, despite achieving a model with high prediction accuracy in the experiment, its practical application was limited. In this study, wolfberry plantation under normal fertilization conditions of local farmers in Ningxia province was selected as the sample collection area. Consequently, the optimal model for predicting LNC of wolfberry tree obtained in this study holds certain practical reference value for the diagnosis of nitrogen nutrients during specific growth stages of

wolfberry (June to August). Additionally, this study combines mathematical transformation with machine learning to provide a spectral selection method and a modeling framework for monitoring wolfberry LNC, which may serve as a useful and valuable tool in nitrogen monitoring and management. To develop accurate, robust and fast model with high reliability, practicability and applicability, the next step should be to validate these models by collecting extensive data for different growth stages, and develop nitrogen prediction models suitable for each growth stages.

Conclusions

Currently, there is limited research on diagnosing nitrogen levels in wolfberry tree using hyperspectral data. In this study, new spectral indices based on transformation spectra were developed and combined with machine learning (ML) to explore the monitoring of wolfberry tree LNC. The study indicates that spectral transformations enhance the reflectance difference in the VIS and SWIR region, highlight the subtle spectral difference in the red-edge region and improve the accuracy of the MLR-NSIs models. However, they do not have a significant effect on the accuracy of the MLR-PVIs models. In the MLR-PVIs models, it is found that ET model constructed by using the original spectra is optimal, with R_c^2, R_v^2 and RPD values of 0.892, 0.576 and 1.510 respectively. In contrast, in the MLR-NSIs models, the RF model constructed by using the first-derivative spectra datasets is optimal for predicting wolfberry LNC, with R_c^2, R_v^2 and RPD values of 0.90, 0.71 and 1.90 respectively. Therefore, although it is feasible to use PVIs to predict wolfberry LNC, the MLR-PVIs models perform worse than the MLR-NSIs model. This study combines spectral transformation with machine learning algorithms to identify a set of NSIs and a machine learning model suitable for detecting wolfberry tree LNC. The set of NSIs are DSIR'$_{(1128,388)}$, DSIR'$_{(745,449)}$ and DSIR'$_{(1945,745)}$ developed by using first-derivative spectral data. This model is a RF model constructed by using this set of NSIs.

Supporting information

S1 Fig. Spectral dataset. The original spectra are derived from the rapid Fourier transform of the raw spectral data.
(XLSX)

S2 Fig. Correlation coefficient. PVIs represents published vegetation indices. Origin-PVIs, reciprocal-PVIs, logarithm-PVIs, square root-PVIs,FD-PVIs, and SD-PVIs represents published vegetation indices calculated from the original spectra, reciprocal spectra, logarithm spectra, square-root spectra, first-derivative spectra and second-derivative spectra datasets, respectively.
(XLSX)

S3 Fig. Determination coefficient, root mean square error and mean absolute error of calibration of the machine learning regression models based on published vegetation indices. R_c^2, RMSE$_c$ and MAE$_c$ represent determination coefficient, root mean square error and mean absolute error of calibration respectively. R, 1/R, logR, $R^{1/2}$, R′ and R″ represent original spectra, reciprocal spectra, logarithmic spectra, square root spectra, first-drivative and second-derivative spectra respectively. AdaBoost, ET and RF represent Adaptive Boosting models, Extra Trees models and Random forest models respectively.
(XLSX)

S4 Fig. Determination coefficient, root mean square error and mean absolute error of validation of the machine learning regression models based on published vegetation indices. R_v^2, RMSE$_v$ and MAE$_v$ represent determination coefficient, root mean square error and mean

absolute error of validation respectively.
(XLSX)

S5 Fig. Determination coefficient, root mean square error and mean absolute error of calibration of the machine learning regression models based on new spectral indices. R_c^2, $RMSE_c$ and MAE_c represent determination coefficient, root mean square error and mean absolute error of calibration respectively.
(XLSX)

S6 Fig. Determination coefficient, root mean square error and mean absolute error of validation of the machine learning regression models based on new spectral indices. R_v^2, $RMSE_v$ and MAE_v represent determination coefficient, root mean square error and mean absolute error of validation respectively.
(XLSX)

S7 Fig. Ratio of percentage deviation of the machine learning regression models. MLR-PVIs models and the MLR-NSIs models are the machine learning regression models based on published vegetation indices and new spectral indices, respectively.
(XLSX)

Acknowledgments

Author Contributions

Data curation: Ligen Zhang.

Supervision: Hao Wang.

Writing – original draft: Yongmei Li.

Writing – review & editing: Hongli Zhao.

References

1. Alon BG, Amnon S, Arnon D, Ran E, Uri Y. Flowering and fruit set of olive trees in response to nitrogen, phosphorus, and potassium. Journal of the American Society for Horticultural Science. 2008; 133: 639–644.

2. Fageria NK, Baligar VC. Enhancing nitrogen use efficiency in crop plants. Advances in Agronomy. 2005; 88: 97–185. http://doi.org/10.1016/S0065-2113(05)88004-6.

3. Zheng HZ, Wei H, Guo SH, Yang X, Feng MX, Jin XQ, et al. Nitrogen and phosphorus co-starvation inhibits anthocyanin synthesis in the callus of grape berry skin. Plant Cell, Tissue and Organ Culture (PCTOC): Journal of Plant Biotechnology. 2020; 142: 313–325. http://doi.org/10.1007/s11240-020-01864-9.

4. Yu CQ, Huang X, Chen H, Godfray HCJ, Wright JS, Hall JW, el at. Managing nitrogen to restore water quality in China. Nature. 2019; 567: 516–520. https://doi.org/10.1038/s41586-019-1001-1 PMID: 30818324

5. Zhang X, Davidson EA, Mauzerall DL, Searchinger TD, Dumas P, Shen Y. Managing nitrogen for sustainable development. Nature. 2015; 528: 51–9. https://doi.org/10.1038/nature15743 PMID: 26595273

6. Li JM, Chen XQ, Yang Q, Shi LS. Deep learning models for estimation of paddy rice leaf nitrogen concentration based on canopy hyperspectral data. Acta Agronomica Sinica. 2021; 47: 1342–1350. https://kns.cnki.net/kcms/detail/11.1809.S.20210104.1140.002.html.

7. Shi ZG, Wei F, Wan R, Li YX, Wang YJ, An W, et al. Impact of nitrogen fertilizer levels on metabolite profiling of the lycium barbarum L.Fruit.Molecules. 2019; 24: 3879. https://doi.org/10.3390/molecules24213879 PMID: 31661883

8. Wang YJ, Li TH, Jin G, Wei YM, Li LQ, Kalkhajeh YK, et al. Qualitative and quantitative diagnosis of nitrogen nutrition of tea plants under field condition using hyperspectral imaging coupled with chemometrics. Journal of the science of food and agriculture. 2020; 100: 161–167. https://doi.org/10.1002/jsfa.10009 PMID: 31471904

9. Ali MM, Al-Ani A, Eamus D, Tan DKY. Leaf nitrogen determination using non-destructive techniques-A review. Journal of Plant Nutrition. 2017; 40: 928–953. http://dio.org/10.1080/01904167.2016.1143954.

10. Pal PK, Singh RD, Prasad R. Non-destructive estimation of chlorophyll and nitrogen content in leaf of Rosa damascena (Mill). Soil Science and Plant Nutrition. 2012; 58: 604–610. http://dio.org/10.1080/00380768.2012.723993.

11. Rubio-Delgado J, Pérez CJ, Rodríguez MA. Predicting leaf nitrogen content in olive trees using hyperspectral data for precision agriculture. Precision Agriculture. 2020; 22: 1–21. http://dio.org/10.1007/s11119-020-09727-1.

12. Camino C, González-Dugo V, Hernández P, Sillero JC, Zarco-Tejada PJ. Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. International Journal of Applied Earth Observation and Geoinformation. 2018; 70: 105–117. http://dio.org/10.1016/j.jag.2018.04.013.

13. Yang ZF, Tian JC, Feng KP, Gong X, Liu JB. Application of a hyperspectral imaging system to quantify leaf-scale chlorophyll, nitrogen and chlorophyll fluorescence parameters in grapevine. Plant Physiology and Biochemistry. 2021; 166: 723–737. https://doi.org/10.1016/j.plaphy.2021.06.015 PMID: 34214782

14. He L, Song X, Feng W, Guo BB, Zhang YS, Wang YH, et al. Improved remote sensing of leaf nitrogen concentration in winter wheat using multi-angular hyperspectral data. Remote Sensing of Environment. 2016; 174: 122–133. http://dio.org/10.1016/j.rse.2015.12.007.

15. Chen PF, Haboudane D, Tremblay N, Wang JH, Vigneault P, Li BG. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. Remote Sensing of Environment. 2010; 114: 1987–1997. http://dio.org/10.1016/j.rse.2010.04.006.

16. Li ZH, Jin XL, Yang GJ, Drummond J, Yang H, Clark B, et al. Remote sensing of leaf and canopy nitrogen status in winter wheat (Triticum aestivum L.) based on N-PROSAIL model. Remote Sensing. 2018; 10: 1463. http://dio.org/10.3390/rs10091463.

17. Wang YN, Li FL, Wang WD, Chen XK, Chang QR.Monitoring of winter wheat nitrogen nutrition based on UAV hyperspectral images. Transactions of the Chinese Society of Agricultural Engineering. 2020; 36: 31–39. http://dio.org/10.11975/j.issn.1002-6819.2020.22.004.

18. Yu FH, Xing SM, Guo ZH, Bai JC, Xu TX. Remote sensing inversion of the nitrogen content in rice leaves using charactertransfer vegetation index. Transactions of the Chinese Society of Agricultural Engineering. 2022; 38: 175–182. http://dio.org/10.11975/j.issn.1002-6819.2022.02.020.

19. Wang JJ, Song XY, Mei X, Yang GJ, Li ZH, Li HL, et al. Sensitive bands selection and nitrogen content monitoring of rice based on Gaussian regression analysis.Spectroscopy and spectral analysis. 2020; 41: 1722–1729. http://dio.org/10.3964/J.ISSN.10000593(2021)06-1722-08.

20. Shu MY, Zhu ZY, Yang XH, Gu XH, Li BG, Ma YT. A spectral decomposition method for estimating the leaf nitrogen status of maize by UAV-based hyperspectral imaging.Computers and Electronics in Agriculture. 2023; 121: 128100. http://dio.org/10.1016/J.COMPAG.2023.108100

21. Milica C, Kang Y, Mladen T, Vito C, Mohamad H, Rossella A, Maria SA. Hyperspectral vegetation indices to assess water and nitrogen status of sweet maize crop. Agronomy. 2022; 12: 2181–2181. http://dio.org/10.3390/AGRONOMY12092181.

22. Lu YL, Zhang XY, Cui YZ, Chao YR, Song GP, Nie CE, et al. Response of different varieties of maize to nitrogen stress and diagnosis of leaf nitrogen using hyperspectral data. Scientific reports. 2023; 13: 5890–5890. https://doi.org/10.1038/s41598-023-31887-z PMID: 37041196

23. Gao LL, Zhu XC, Li C, Cheng LZ. Evaluation of the nitrogen content during the new-shoot-growing stage in apple leaves using two-dimensional correlation spectroscopy. PloS one. 2017; 12: 0186751. https://doi.org/10.1371/journal.pone.0186751 PMID: 29073247

24. Fan K, Li FL, Chen XK, Li ZF, Mulla DJ. Nitrogen balance index prediction of winter wheat by canopy hyperspectral transformation and machine learning. Remote Sensing. 2022; 14: 3504–3504. http://dio.org/10.3390/RS14143504.

25. Li FL, Wang L, Liu J, Wang YN, Chang QR. Evaluation of leaf N concentration in winter wheat based on discrete wavelet transform analysis. Remote Sensing. 2019; 11: 1331. http://dio.org/10.3390/rs11111331.

26. Guo JB, Zhang JJ, Xiong SP, Zhang ZY, Wei QQ, Zhang W, et al. Hyperspectral assessment of leaf nitrogen accumulation for winter wheat using different regression modeling. Precision Agriculture. 2021; 22: 1–25. http://dio.org/10.1007/S11119-021-09804-Z.

27. Li YM, Chen XD, Li F, Zhang XJ. Effect of formula fertilization on yield and quality of lycium barbarum based on fertigation. Northern horticulture. 2018; 22: 161–168. http://dio.org/10.11937/bfyy.20180801.

28. Ma ZH, Yin J, Yang YP, Sun FB, Yang Z. Effect of water and nitrogen coupling regulation on the growth, physiology,yield, and quality attributes and comprehensive evaluation of wolfberry (Lycium barbarum L.). Frontiers in plant science. 2023; 14: 1130109–1130109. https://doi.org/10.3389/fpls.2023.1130109 PMID: 37416888

29. Zhao AX, Tang XJ, Zhang ZH, Liu JH. Optimizing Savitzky-Golay Parameters and Its Smoothing Pre-treatment for FTIR Gas Spectra. Spectroscopy and Spectral Analysis. 2016; 36: 1340–1344. http://dio.org/10.3964/j.issn.1000-0593(2016)05-1340-05. PMID: 30001002

30. Bi YM, Yuan KL, Xiao WQ, Wu JZ, Shi CY, Xia J, et al. A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. Analytica Chimica Acta. 2016; 909: 30–40. https://doi.org/10.1016/j.aca.2016.01.010 PMID: 26851082

31. Savitzky A. Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry. 2002; 36: 1627–1639. http://dio.org/10.1021/ac60214a047.

32. Xu P, Cao Y, Xue LY, Zhang JC, Zhu L, Chen BQ, et al. Study on sparse representation and measurement matrices of compressive sensing of plant hyperspectral data for retrieving plant physiological and biochemical parameters. Biosystems Engineering. 2019; 182: 38–53. http://dio.org/ 10.1016/j.biosystemseng.2019.02.011.

33. Wen SY, Shi N, Lu JW, Gao QW, Hu WR, Cao ZDY, et al. Continuous wavelet transform and back propagation neural network for condition monitoring chlorophyll fluorescence parameters Fv/Fm of rice reaves. agriculture. 2022; 12: 1197–1197. http://dio.org/10.3390/AGRICULTURE12081197.

34. He T. Study on Land Quality Hyperspectral Remote Sensing Monitoring Method. 2023; Wuhan University, Wuhan. Doctoral thesis.

35. Huang MX, Wang K, Shi Z, Gong JH, Li HY, Chen JL. Quantitative evaluation of soil hyperspectra denoising with different filters. Spectroscopy and Spectral Analysis. 2009; 2: 722–725. http://dio.org/10.3964/j.issn.1000-0593(2009)03-0722-04.

36. Duan DD, He YB, Luo SJ, Wang ZZ. Analysis on the ability of distinguishing potato varieties with different hyperspectral parameters. Spectroscopy and Spectral Analysis. 2018; 38: 3215–3220. http://dio.org/10.3964/j.issn.1000-0593(2018)10-3215-06.

37. Tian AH, Xiong HG, Zhao JS, Fu CB.Mechanism improvement for pretreatment accuracy of field spectra of saline soil using fractional differential algorithm. Spectroscopy and Spectral Analysis. 2019; 39: 2495–2500. http://dio.org/10.3964/j.issn.1000-0593(2019)08-2495-06.

38. Moros J. Vibrational spectroscopy provides a green tool for multi-component analysis. Trends in Analytical Chemistry. 2009; 29: 578–591. http://dio.org/10.1016/j.trac.2009.12.012.

39. Luo SJ, He YB, Li Q, Jiao WH, Zhu YQ, Yu JK, et al. Assessment of unified models for estimating potato leaf area index under water stress conditions across ground-based hyperspectral data. Journal of Applied Remote Sensing. 2020; 14: 014517. http://dio.org/10.1117/1.JRS.14.014517.

40. Pearson R L, Miller L D. Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. Remote Sensing of Environment. 1972; 1: 1355.

41. Rouse J W, Haas R H, Schell J A. Monitoring the vernal advancements and retro gradation of natural vegetation. NASA/GSFC, Final Report, Greenbelt, MD, USA. 1974; 1–137.

42. Vogelmann J E, Rock BN, Moss DM. Red edge spectral measurements from sugar maple leaves. International Journal of Remote Sensing. 1993; 14: 1563–1575.1. http://dio.org/10.1080/01431169308953986.

43. Zarco-Tejada PJ, Pushnik JC, Dobrowski S, Ustin SL. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. Remote Sensing of Environment. 2003; 84: 283–294. http://dio.org/10.1016/S0034-4257(02)00113-X.

44. Dash J, Curran PJ. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). Advances in Space Research. 2007; 39: 100–104. http://dio.org/10.1016/j.asr.2006.02.034.

45. Gitelson AA, Merzlyak MN.Remote estimation of chlorophyll content in higher plant leaves.International Journal of Remote Sensing. 1997; 18: 2691–2697. http://dio.org/ 10.1080/014311697217558.

46. Gitelson AA, Merzlyak MN. Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L.and Acer platanoides L.leaves.Spectral features and relation to chlorophyll estimation. Journal of Plant Physiology. 1994; 143: 286–292. http://dio.org/10.1016/S0176-1617(11)81633-0.

47. Sims DA, Gamon JA. Relationships between leaf pigment content and spectral reflectance across a wide range of species,leaf structures and developmental stages. Remote Sensing of Environment. 2002; 81: 337–354.1. http://dio.org/10.1016/S0034-4257(02)00010-X.

48. Gitelson AA, Viña A, Ciganda V, Rundquist DC, Arkebauer TJ. Remote estimation of canopy chlorophyll content in crops. Geophysical Research Letters. 2005; 32: 403-1–403-4. http://dio.org/10.1029/2005GL022688.

49. Hansen PM, Schjoerring JK. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote sensing of environment. 2003; 86: 542–553. http://dio.org/10.1016/S0034-4257(03)00131-7.

50. Reyniers M, Walvoort DJJ, Baardemaaker JD. A linear model to predict with a multi-spectral radiometer the amount of nitrogen in winter wheat. International Journal of Remote Sensing. 2006; 27: 4159–4179. http://dio.org/10.1080/01431160600791650.

51. Barnes EM, Clarke TR, Richards SE. Coincident detection of crop water stress,nitrogen status and canopy density using ground based multispectral data.Proceedings of the 5th International Conference on Precision Agriculture. Bloomington, MN. 2000; 16–19.

52. Daughtry CST, Walthall CL, Kim MS, Colstoun B, McMurtrey E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote sensing of Environment. 2000; 74: 229–239. http://dio.org/10.1016/S0034-4257(00)00113-9.

53. Haboudane D, Miller JR, Pattey E, Zarco-Tejada PJ, Strachan IB. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies:Modeling and validation in the context of precision agriculture. Remote Sensing of Environment. 2004; 90: 337–352. http://dio.org/10.1016/j.rse.2003.12.013.

54. Eitel JUH, Vierling LA, Litvak ME, Long DS, Schulthess U, Ager AA, et al. Red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. Remote Sensing of Environment. 2011; 115: 3640–3646. http://dio.org/10.1016/j.rse.2011.09.002.

55. Metternicht G. Vegetation indices derived from high-resolution airborne videography for precision crop management. International Journal of Remote Sensing. 2003; 24: 2855–2877. http://dio.org/10.1080/01431160210163074.

56. Garbulsky MF, Peñuelas J, Gamon J, Inoue Y, Filella I. The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies. Remote Sensing of Environment. 2010; 115: 281–297. http://dio.org/10.1016/j.rse.2010.08.023.

57. Gitelson AA, Merzlyak MN. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. Journal of plant physiology. 1996; 148: 494–500. http://dio.org/ 10.1016/S0176-1617(96)80284-7.

58. Peñuelas J, Gamon JA, Fredeen AL, Merino J, Field CB. Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. Remote Sensing of Environment. 1994; 48: 135–146. http://dio.org/10.1016/0034-4257(94)90136-8.

59. Schleicher TD, Bausch WC, Delgado JA. Evaluation and refinement of the nitrogen reflectance index (NRI) for site-specific fertilizer management. 2001 ASAE Annual International Meeting, St-Joseph, MI, USA. ASAE Paper. 2001(01–11151).

60. Peñuelas J, Baret F, Filella I. Semi-empirical indices to assess cartotenoids/chlorophylla ratio from leaf spectral reflectance. Photosynthetica. 1995; 31: 221–230.

61. Broge NH, Leblanc E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sensing of Environment. 2001; 76: 156–172. http://dio.org/10.1016/S0034-4257(00)00197-8.

62. Gitelson AA, Kaufman YJ, Stark R, Rundquist D. Novel algorithms for remote estimation of vegetation fraction.Remote Sensing of Environment. 2002; 80: 76–87. http://dio.org/10.1016/S0034-4257(01)00289-9.

63. Hunt ER Jr, Daughtry CST, Eitel JUH, Long DS. Remote sensing leaf chlorophyll content using a visible band index. Agronomy Journal. 2011; 103: 1090–1099. http://dio.org/10.2134/agronj2010.0395.

64. Haboudane D, Miller JR, Tremblay N, Zarco-Tejada PJ, Dextraze. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sensing of Environment.2002; 81: 416–426. http://dio.org/ 10.1016/S0034-4257(02)00018-4.

65. Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices. Remote Sensing of Environment. 1996; 55: 95–107. http://dio.org/10.1016/0034-4257(95)00186-7.

66. Francisca López Granados, José M . Peña Barragán, Luis García Torres, M. Teresa Gómez Casero, Montserrat Jurado Expósito, Ricardo Fernández Escobar. Assessing nitrogen and potassium deficiencies in olive orchards through discriminant analysis of hyperspectral data. Joural of the American Society for Horticultural Science. 2007; 132: 611–618. http://dio.org/10.21273/JASHS.132.5.611.

67. Srivastava PK, Gupta M, Singh U, Prasad R, Pandey PC, Raghubanshi AS, et al. Sensitivity analysis of artificial neural network for chlorophyll prediction using hyperspectral data. Environment, Devel opment and Sustainability. 2020; 23: 1–16. http://dio.org/10.1007/s10668-020-00827-6.

68. Huang Z, Turner BJ, Dury SJ, Wallis IR, Foley W.J. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. Remote Sensing of Environment. 2004; 93: 18–29. http://dio.org/10.1016/j.rse.2004.06.008.

69. Onan A. Mining opinions from instructor evaluation reviews: A deep learning approach.Computer Application in Engineering Education. 2020; 28: 117–138. https://doi.org/10.1002/ cae.22179.

70. Buyrukoğlu G, Buyrukoğlu S, Topalcengiz Z. Comparing regression models with count data to artificial neural network and ensemble models for prediction of generic escherichia coli population in agricultural ponds based on weather station measurements. Microbial Risk Analysis. 2021; 19: 100171. https://doi. org/10.1016/J.MRAN.2021.100171.

71. Yu RY, Zhu XC, Cao SJ, Xiong JL, Wen X, Jiang YM, et al. Estimation of chlorophyll content in apple leaves based on imaging spectroscopy. Journal of Applied Spectroscopy. 2019; 86: 457–464. http://dio.org/10.1007/s10812-019-00841-1.

72. Majeed HM, Khalid AM, Faidhalrahman K, Nadhir AA. An extra tree regression model for discharge coefficient prediction:novel, practical applications in the hydraulic sector and future research directions. Mathematical Problems in Engineering. 2021; 2021: 19. http://dio.org/10.1155/2021/7001710.

73. Yi CH, Sogah Dk, Boyce M, Degterev A, Christofferson DE, Yuan J. A genome-wide RNAi screen reveals multiple regulators of caspase activation. The Journal of cell biology. 2007; 179: 619–26. https://doi.org/10.1083/jcb.200708090 PMID: 17998402

74. Pohl C, Van Genderen J. Remote sensing image fusion:A practical guide; CRC Press: Boca Raton, FL, USA, 2016.

75. Williams PC. Variables affecting near-infrared reflectance spectro-scopic analysis.In Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists. 1987; 143–166.

76. Shi T, Cui L, Wang J, Fei T, Chen Y, Wu G. Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. Plant and Soil. 2013; 366: 363–375.

77. Gaston E, Frías Jesús M, Cullen Patrick J, O'Donnell CP, Gowen Aoife A. Prediction of polyphenol oxidase activity using visible near-infrared hyperspectral imaging on mushroom (*Agaricus bisporus)* caps. Journal of agricultural and food chemistry. 2010; 58: 6226–33. http://dio.org/10.1021/jf100501q.

78. Im J, Jensen JR. Hyperspectral remote sensing of vegetation. Geography Compass. 2008; 2: 1943–1961. http://dio.org/10.1111/j.1749-8198.2008.00182.x.

79. Zhang J, Xu B, Feng HK, Jing X, Wang JJ, Ming SK, et al. Monitoring nitrogen nutrition and grain protein content of rice based on ensemble learning. Spectroscopy and spectral analysis. 2022; 42: 1956–1964. http://dio.org/10.3964/j.issn.1000-0593(2022)06-1956-09.

80. Zheng YM, Zhang TQ, Zhang J, Chen XD, Shen XG. Influence of Smooth,1st Derivative and Baseline Correction on the Near-Infrared Spectrum Analysis with PLS. Spectroscopy and Spectral Analysis. 2004; 24: 1546–1548. https://doi.org/1000-0593(2004)12-1546-03. PMID: 15828323

81. Tian GY, Yuan HF, Chu XL, Liu HY, Lu WZ. Near infrared spectra (NIR) analysis of octane mumber by wavelet denoising-derivative method.Spectroscopy and Spectral Analysis. 2005; 25: 516–520. https://doi.org/1000-0593 (2005) 04-0516-05.

82. Curran PJ. Remote sensing of foliar chemistry.Remote Sensing of foliar chemistry. 1989; 30: 271–278. http://dio.org/10.1016/0034-4257(89)90069-2.

83. Zarco-Tejada PJ, Miller JR, Noland TL, Mohammed GH, Sampson PH. 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. 2001; 39: 1491–1507. http://dio.org/10.1109/36.934080.

84. Wu CY, Niu Z, Tang Q, Huang WJ. Estimating chlorophyll content from hyperspectral vegetation indices:Modeling and validation. Agricultural and Forest Meteorology. 2008; 148: 1230–1241. http://dio.org/10.1016/j.agrformet.2008.03.005.

85. Yang GJ, Zhao CJ, Li ZH. Quantitative remote sensing of crop nitrogen nutrition and its application. Beijing:Science Press; 2019.

86. Ceccato P, Gobron N, Flasse S, Pinty B, Tarantola S.Designing a spectral index to estimate vegetation water content from remote sensing data:part 1. Remote Sensing of Environment. 2002; 82: 188–197. http://dio.org/10.1016/S0034-4257(02)00037-8.

87. Tian AH, Xiong HG, Zhao JS, Fu CB. Mechanism improvement for pretreatment accuracy of field spectra of saline soil using fractional differential algorithm. Spectroscopy and Spectral Analysis. 2019; 39: 2495–2500. http://dio.org/10.3964/j.issn.1000-0593(2019)08-2495-06.

88. Kokaly R, Clark RN. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sensing. 1999; 67: 267–287.

89. Yao X, Huang Y, Shang GY, Zhou C, Cheng T, Tian YC, et al. Evaluation of Six Algorithms to Monitor Wheat Leaf Nitrogen Concentration. Remote Sensing. 2015; 7: 14939–14966. http://dio.org/10.3390/rs71114939.

90. Gnyp ML, Miao Y, Yuan F, Ustin SL, Yu K,Yao Y,et al. Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. Field Crops Research.2014; 155:42–55. http://dio.org/10.1016/j.fcr.2013.09.023.

91. Li D, Wang CY, Liu W, Peng ZP, Huang SY, Huang JC, et al. Estimation of litchi (Litchi chinensis Sonn.) leaf nitrogen content at different growth stages using canopy reflectance spectra. European Journal of Agronomy. 2016; 80: 182–194. http://dio.org/10.1016/j.eja.2016.08.001.