Combining the fractional order derivative and machine learning for leaf water content estimation of spring wheat using hyper-spectral indices

Zinhar Zununjan[1], Mardan Aghabey Turghan[3], Mutallip Sattar[4], Nijat Kasim[1*], Bilal Emin[1] and Abdugheni Abliz[2]

Abstract

Leaf water content (LWC) is a vital indicator of crop growth and development. While visible and near-infrared (VIS–NIR) spectroscopy makes it possible to estimate crop leaf moisture, spectral preprocessing and multiband spectral indices have important significance in the quantitative analysis of LWC. In this work, the fractional order derivative (FOD) was used for leaf spectral processing, and multiband spectral indices were constructed based on the band-optimization algorithm. Eventually, an integrated index, namely, the multiband spectral index (MBSI) and moisture index (MI), is proposed to estimate the LWC in spring wheat around Fu-Kang City, Xinjiang, China. The MBSIs for LWC were calculated from two types of spectral data: raw reflectance (RR) and the spectrum based on FOD. The LWC was estimated by combining machine learning (K-nearest neighbor, KNN; support vector machine, SVM; and artificial neural network, ANN). The results showed that the fractional derivative pretreatment of spectral data enhances the implied information of the spectrum (the maximum correlation coefficient appeared using a 0.8-order differential) and increases the number of sensitive bands, especially in the near-infrared bands (700–1100 nm). The correlations between LWC and the two-band index ($RVI_{1156, 1628\,nm}$), three-band indices ($3BI\text{-}3_{(766, 478, 1042\,nm)}$, $3BI\text{-}4_{(1129, 1175, 471\,nm)}$, $3BI\text{-}5_{(814, 929, 525\,nm)}$, $3BI\text{-}6_{(1156, 1214, 802\,nm)}$, $3BI\text{-}7_{(929, 851, 446\,nm)}$) based on FOD were higher than that of moisture indices and single-band spectrum, with r of -0.71^{**}, 0.74^{**}, 0.73^{**}, -0.72^{**}, 0.75^{**} and -0.76^{**} for the correlation. The prediction accuracy of the two-band spectral indices ($DVI_{(698, 1274\,nm)}$ $DVI_{(698, 1274\,nm)}$ $DVI_{(698, 1274\,nm)}$) was higher than that of the moisture spectral index, with R^2 of 0.81 and R^2 of 0.79 for the calibration and validation, respectively. Due to a large amount of spectral indices, the correlation coefficient method was used to select the characteristic spectral index from full three-band indices. Among twenty seven models, the $FWBI\text{-}3BI_{\_0.8\,order}$ model performed the best predictive ability (with an R^2 of 0.86, RMSE of 2.11%, and RPD of 2.65). These findings confirm that combining spectral index optimization with machine learning is a highly effective method for inverting the leaf water content in spring wheat.

Keywords Spring wheat, Leaf water content, Multiband Spectral Index, Fractional order derivative, Machine learning

*Correspondence:
Nijat Kasim
Nejatkasim@126.com
Full list of author information is available at the end of the article

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 2 of 24

Introduction

Water is the main medium for transporting mineral nutrients and conducting physiological and biochemical reactions in plants [1–3]. Lack of water can lead to weakened plant transpiration, hindered mineral transport, and decreased chlorophyll content, and ultimately, it restricts the accumulation of assimilates [4, 5]. Spring wheat can be divided into three stages: early stage (emergence and tillering), middle stage (jointing and booting), and late stage (heading and maturing) [6, 7]. During the heading stage, wheat grains begin to develop rapidly, absorbing nutrients and forming components such as starch and protein [8]. It is necessary to enhance irrigation, maintain soil moisture, and apply fertilizer in a timely manner to promote wheat growth [9]. The water content in the internal tissues of leaves can reflect information about soil moisture, crop growth and development, disease resistance, and other factors [10, 11], making it an important indicator for evaluating crop growth and development [12]. Therefore, obtaining real-time and accurate information on leaf water content (LWC) in field crops has important guiding significance for assisting regional agricultural production management.

With the development of remote sensing techniques, remotely sensed data have been widely used to accurately and non-destructively monitor crop parameters [13, 14]. Hyperspectral had the characteristics of multiple bands, strong continuity, and large information; and was performed accurately and quickly with the following objectives: diagnosis of crop leaf water, nutrient status, monitoring of crop growth, spatial variation information of crop biochemical components and evaluation of crop yield [15, 16]. Hyperspectral remote sensing-based modeling can utilize visible and near-infrared reflectance (Vis–NIR) spectra to rapidly respond to changes in crop parameters [17]. However, Hyperspectral data are characterized by a large amount of data and multicollinearity and are usually composed of three types of spectral information: valid information, redundant information, and invalid information [18, 19]; and spectral preprocessing constitutes an important step in spectral modeling analysis. Several spectral preprocessing methods, including spectral continuum removal, spectral logarithmic, first/second-order derivative, and reciprocal logarithmic, and to improve prediction of crop parameters [20]. In general, integer differentials (first/second-order derivative) are mainly used to reduce the influence of baseline drift and isolate overlapping peaks. However, the integer differentials fail to account for some subtle details regarding the reflectance spectra [21, 22]. In contrast, fractional-order derivatives (FOD) not only refine the spectral spacing and amplify weak spectral absorption characteristics in a small interval but also reflect changes

in spectral information to some extent [74, 75]. Furthermore, several studies analyzed the relationship between crop components and hyperspectral data based on integer differentials, identified spectral absorption regions (approximately 970 nm, 1200 nm, 1450 nm, 1940 nm, and 2500 nm) of crop leaf moisture, followed by the selection of sensitive bands to predict LWC using partial least square regression (PLSR), Support vector regression (SVR), linear or non-linear functions, and their accuracy is validated [23, 24]. However, there is a lack of accurate research on the simulation models of LWC based on FOD.

The use of spectral indices is a simple and effective method for measuring surface properties, and the band optimization algorithm is widely used in the development of hyperspectral techniques [12, 49]. Compared to single sensitive spectral data, this method has the ability to provide more spectral features and enhance the relationship between crop parameters and spectral features [25]. The spectral index constructed by combining two sensitive bands can improve the accuracy of crop parameter estimation [26, 27]. For example, the ratio of $R_{1600\ nm}/R_{820\ nm}$ was significantly correlated with the effective moisture thickness of leaves [28, 29], while the application of $R_{1450\ nm}/R_{1940\ nm}$ can provide a better estimate of plant water status [30]. The depth and area of the absorption valley in the spectral curve are more sensitive to the water content of the crop leaves, and there is a linear positive correlation between the characteristic absorption depth and the area of water near 1450 nm [31, 32]. $DR_{1647\ nm}/DR_{1133\ nm}$ and $DR_{1653\ nm}/DR_{1687\ nm}$ are based on derivative spectra and have higher fitting accuracy with crop leaf water compared with that of single band indices [33]. It was demonstrated that combining two bands within the 400–2400 nm range to construct spectral indices (such as RVI, NDVI, and DVI) can improve sensitivity to moisture levels in wheat leaves and enhance the modeling accuracy [47, 49]. However, in practice, the band determination of the sensitivity index depends on the two-dimensional contour map, so the band optimization algorithm is mostly limited to the two-band index form [48]. In previous studies, few researchers have considered extending the spectral index method to more than two bands, particularly optimizing three-band indices and estimating crop leaf moisture by combining traditional spectral indices. Therefore, there is a need to further explore the integrated index based on the fusion of multi-band spectral indices.

Remote sensing technology and its integration with machine learning are examples of effective and low-cost acquisition, particularly in the field of environmental studies and earth sciences for optimal management [69]. Several studies have been conducted on hyperspectral

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 3 of 24

remote sensing and inversion models. For instance, Felegari et al., mapped the Cd concentration and introduced the most suitable regression models, including support vector regression (SVR), partial least square regression (PLSR), and artificial neural networks (ANN) [70]. A Wavelet-Attention convolutional neural network (WA-CNN), Random forest (RF) and support vector machine (SVM) algorithms were utilized to automatically map the crops over the agricultural lands [71]. Seyed et al., proposed the two architectures: the first model includes 2D-CNN, skip connections, and LSTM-Attentions and the second model comprises 3D-CNN, skip connections, and Conv-LSTM Attention (The Input data given from MODIS products) [72]. Many strategies have been used to evaluate agricultural products, such as Deep-Yield, CNN-LSTM, and Conv-LSTM. And it showed that machine learning is widely applied in remote sensing technology. However, In the above-mentioned research, there are more studies on the application of machine learning in multi-source remote sensing image, while there are fewer applications of high-precision modeling in ground-based hyperspectral.

The LWC and hyperspectral data were acquired on heading stage of spring wheat. The hypothesis of this study was that band-optimization algorithm using fractional order derivative (FOD) was still effective, and thus, the indices (two-band index and three-band index) calculated based on well processing spectrum will better predict LWC. The combined indices (spectral, two-band index and three-band index) will help to improve the accuracy of predictions of LWC values with the machine learning approaches at heading stage of spring wheat [31, 34].

In this study, twenty seven different models were established separately based on multiband spectral data (single bands, moisture indices, two-band spectral indices and three-band spectral indices) and machine learning (KNN, ANN and SVR). The main purpose of this investigation was to provide a future reference for hyperspectral monitoring of spring wheat leaf moisture under similar production conditions. To achieve this purpose, the major sub-goals were defined:

(1) generate and analyze the spectral indices with FOD (with an interval of 0.2, ranging from 0 to 2) for LWC of spring wheat;

(2) to evaluate the performance of newly developed spectral indices and combination spectral indices (spectral, moisture indices, two-band index and three-band index) based on FOD for LWC at heading stage of spring wheat;

(3) to compare the ability of twenty seven models established by the machine learning approaches to monitor LWC and to identify the optimal model among them.

Materials and methods

Location of the study area

The study was carried out in Fu-kang City, Xinjiang China, situated at an altitude of 577 m and characterized by a temperate continental dry climate, the study sites are depicted in Fig. 1. The experimental site is a typical arid farming area, and the site experiences an average annual temperature of 6.7 °C and an annual precipitation of 205 mm. During the summer season, temperatures are exceptionally high, with significant diurnal variations [34]. In the experimental site, sowing of spring wheat commenced on April 20, 2017 (day of the year (DOY) 96), utilizing 225 kg of seeds per hectare and 17 cm row spacing. Diammonium phosphate (150 kg/ha), ammonium sulfate (150 kg/ha), and potash fertilizer (105 kg/ha) were applied during sowing. At the three-leaf stage, fertilized urea (20 kg) was applied, along with drip irrigation and 750 to 900 m^3/ha of irrigation in response to rainfall. Chemical methods were employed to remove weeds from the field, 100 ml of 20% Bromoxynil octanoate E.C. was sprayed before jointing stage of the spring wheat (from the 2-leaf stage) [34].

Field data collection and LWC determination

To ensure consistency in the measurements, the same treatments were maintained throughout the growing season, and a drip irrigation system was used to determine the timing and amount of water applied. The research area consists of 154 sampling plots, the size of each measuring plot is 1 m × 1 m. Measurements were taken at nine distinct points within each plot (as illustrated in Fig. 2a). Daily average precipitation, maximum and minimum air temperature were collected in the field over the course of a year (as illustrated in Fig. 2b). In 2017, the highest daily average precipitation occurred in July, reaching 38 mm. The highest daily temperature occurred in June and August, while the lowest temperature occurred in January. The lowest temperature in June was 11 °C.

We obtained plant samples (on June 04, 2017) from each small sampling area in sealed plastic bags and ensured that plant moisture was not lost. In the laboratory, the leaves and stems of spring wheat were separated, and the weight of each sample was measured with an electronic balance with a sensitivity of 0.0001 g and recorded as the fresh weight [35]. Then, all the samples were dried in an oven (80 °C) for 36 h and weighed, and the dry weight was recorded. The 154 samples were randomly divided into two groups: one was a modeling sample (70%), which was used to establish a spring wheat leaf moisture content prediction model; the other was a verification sample (30%), which
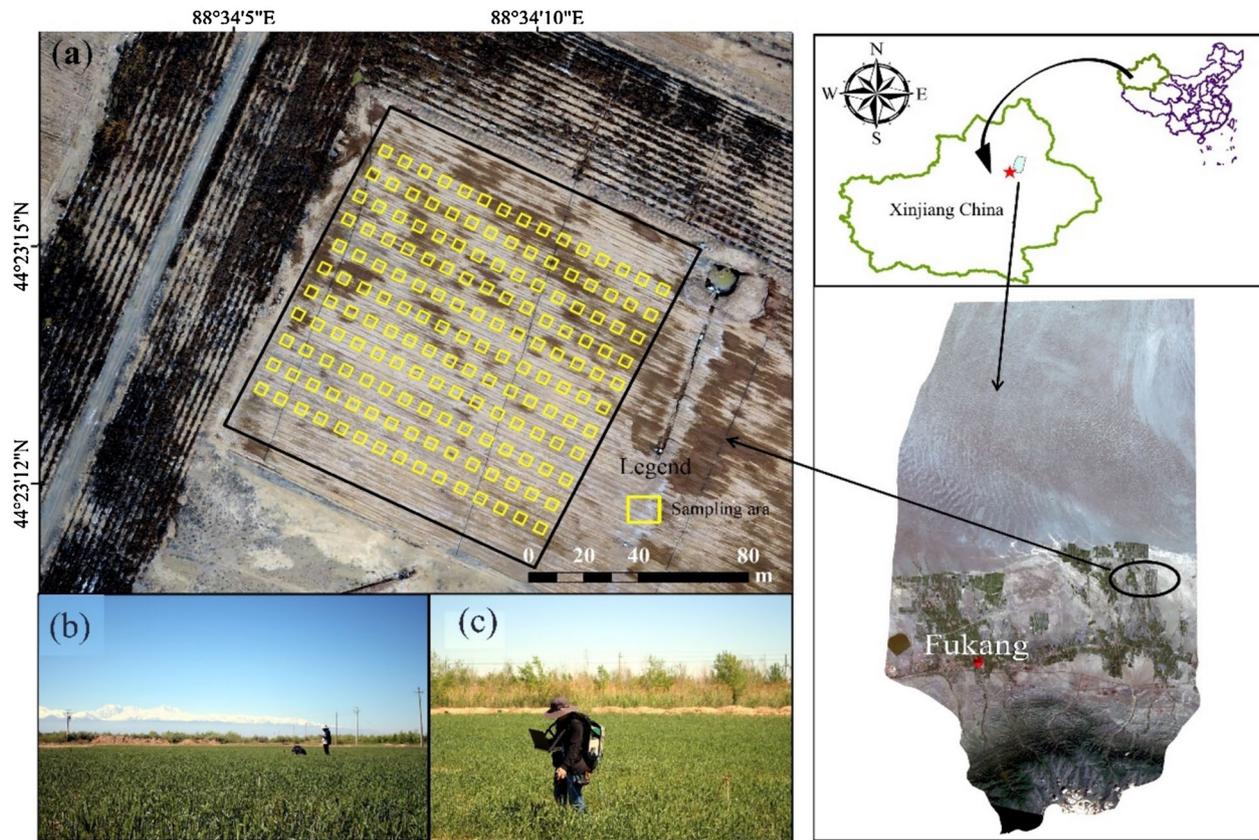
Fig. 1 Geographical location of the study area and planting region for spring wheat. **a** Sampling area for field data. **b**, **c** Was spectral data collection
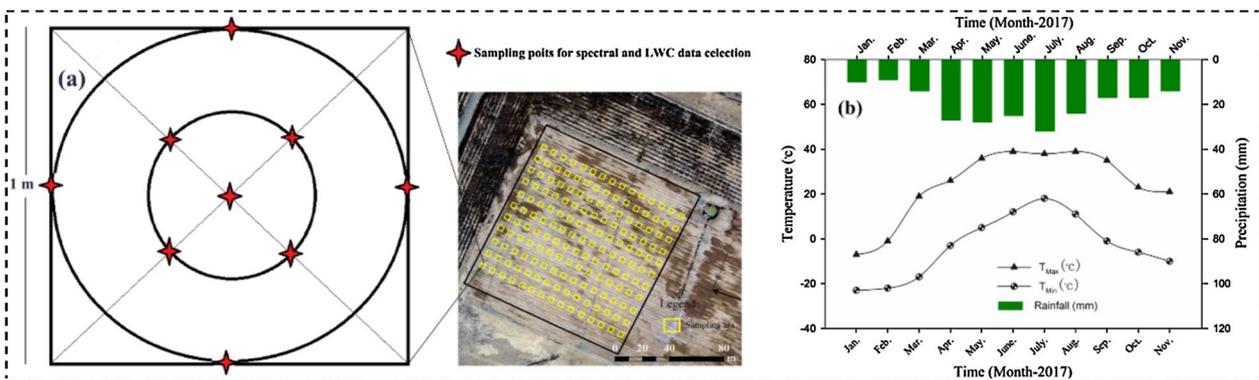
Fig. 2 Sample design and climate change. **a** Design of sampling points for field data in the study area. **b** For the 1-year study, daily average precipitation, maximum and minimum air temperature, and precipitation data were collected

was used to verify the established estimation model [36]. The leaf water content was calculated according to formula (1).

$$LWC = \frac{FW - DW}{FW} \times 100\%,$$ (1)

where FW and DW are the fresh and dry weight (g) of spring wheat leaves, respectively.

Hyperspectral measurement and preprocessing

Spectral reflectance data were collected using the American ASD FieldSpec3 spectrometer, with a band range of

350–2500 nm [37]. The sampling interval was 1.4 nm for the range of 350–1000 nm and 2 nm for the range of 1000–2500 nm [38]. Data collection was carried out in cloudless and sunny weather, the collection time was 10:00 am–02:00 pm, and whiteboard correction was performed every 3–5 min [34]. Ten spectral curves were collected for each sample with a measurement interval of 0.1 s. The average value of these spectral curves was used as the spectral data for that particular sample.

The measured spectral data are preprocessed by removing the noisy edge bands (350–399 nm and 2401–2500 nm) as well as the infrared bands (1355–1444 nm and 1777–1949 nm). Then, the remaining spectral curves are smoothed using mathematic morphological filtering [39], as shown in Fig. 3. Finally, the processed spectral data are used as the basis for calculating the spectral indices in the subsequent step.

Fractional order derivative

The concept of fractional order derivative (FOD) is an extension of the idea of integer order derivatives. This field is dedicated to studying the properties and

applications of derivatives with arbitrary orders [40, 41], as noted by Ortigueira et al. in 2011 and Pan et al. in 2013. The FOD method has been successfully employed in system modeling, signal filtering, and pattern recognition. There are three main types of FOD algorithms: Riemann–Liouville (R–L), Grünwald–Letnikov (G–L), and Caputo [42], as described by Saadia et al. in 2016.

The Grünwald–Letnikov (G–L) definition is relatively simple and was used in our study. In general, the first derivative of a function is defined as follows:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{2}$$

where h is the increment of the independent variable x. Then the second derivative of function can be defined as follow:

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}$$
$$= \lim_{h \to 0} \frac{f(x+2h) - f2(x+h) + f(x)}{h^2}. \tag{3}$$
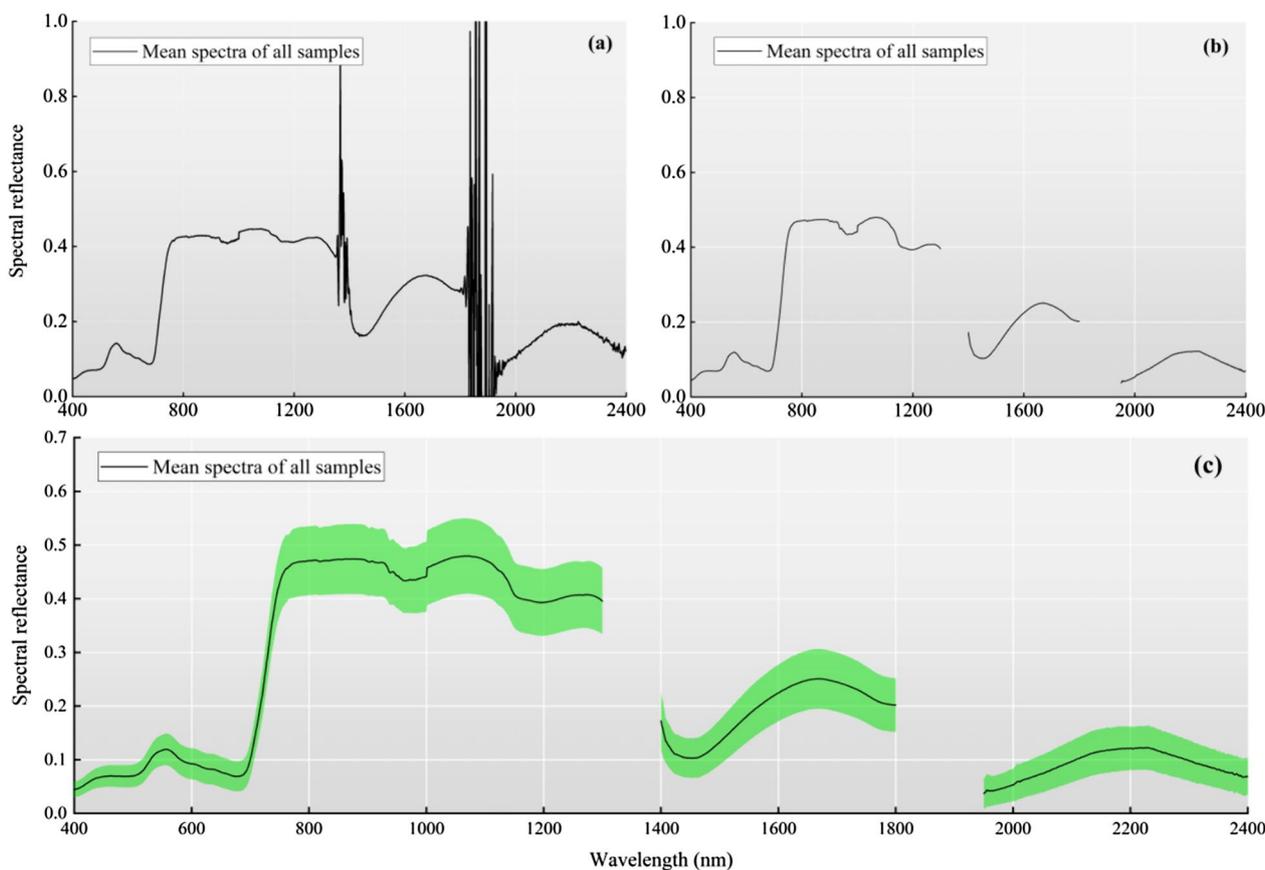
Fig. 3 Preprocessing steps for hyperspectral data. **a** Mean raw spectra of all spring wheat leaf samples (n = 154) and (**b**) removal of the external disturbance spectra of all spring wheat samples. **c** The green shaded areas represent the standard deviations of the spectra

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 6 of 24

If the derivative order of the function $f(x)$ is increased to the higher order (v), then the vth derivative order of the function $f(x)$ can be expressed as:

$$f^v(x) = \lim_{h \to 0} \frac{1}{h^v} \sum_{m=0}^{v} (-1)^m \binom{v}{m} f(x - mh). \qquad (4)$$

By substituting the Gamma function into the binomial coefficient and extending the fractional order to non-integer orders, we can obtain the G-L formula for the v-order fractional derivative in the interval [a, b].

$$f^v(x) = \lim_{h \to 0} \frac{1}{h^v} \sum_{m=0}^{[(b-a/h)]} (-1)^m \frac{\Gamma(v+1)}{m!\Gamma(v-m+1)} f(x - mh), \qquad (5)$$

where h is the step length and is set to 1 in this study, and $[(b-a/h)]$ is the integer part of $(b-a/h)$. Which can be converted to:

$$\frac{d^v f(x)}{dx^v} \approx f(x) + (-v)f(x-1) + \frac{(-v)(-v+1)}{2} f(x-2) + \cdots \frac{\Gamma(-v+1)}{m!\Gamma(-v+m+1)} f(x-m). \qquad (6)$$

To develop spectra-FOD, Eq. (5) was implemented using MATLAB R2014a (The MathWorks Inc.: Natick, MA, USA). The value of v was set to values between 0 and 2 in increments of 0.2 at each step. It is noteworthy that $v=0$ indicated that the raw reflectance was not processed.

Spectral index selection

The spectral index comprises several narrow or wide bands combined through some mathematical transformation, which not only considers the interaction between bands but also improves the response to the measured attributes to some extent. Previous studies have mostly used two-band spectral indices for environmental modeling and attribute quantification. Based on previous studies, two-band water vegetation indices sensitive to water content in crop leaves were collected. The 12 spectral indices selected in this study are listed in Table 1.

The 12 moisture indices listed in Table 1 are commonly used and are sensitive spectral indices for assessing the water content in crop leaves. These spectral indices have fixed bands and serve as the reference for comparing and analyzing the performance of optimized band combinations in this study. The main focus is on assessing the sensitivity strength between traditional spectral indices and band-optimized spectral indices.

Recent studies have calculated the correlation coefficients between two given bands (λ_1 and λ_2) in the Vis–NIR range and the attributes to be measured and have displayed the results visually [48, 49]. This two-dimensional correlation analysis method is beneficial for visualizing the external response and internal meaning of spectra. For example, the normalized difference vegetation index (NDVI) places the strongest reflection band and the weakest reflection band in the numerator and denominator, respectively, and further enlarges the gap between the bands by a normalized ratio operation to maximize the sensitivity of the attributes of the objects to be measured [50, 51].

The addition of a third band in a specific sensitive region to the two-band spectral index can often improve the accuracy of the index estimation, enhance anti-interference ability, and eliminate the saturation phenomenon of commonly used two-band indices [52, 53]. In this study, we used the entire dataset and band optimization algorithm to determine the best wavelength combination.

Table 1 Traditional two-band combined spectral index

Moisture index	Abbreviation	Equation	References
Water index	WI	$R_{900\,nm}/R_{970\,nm}$	[43]
Water band index	WBI	$R_{970\,nm}/R_{900\,nm}$	[44]
Floating position water band index	FWBI	$R_{900\,nm}/Min(R_{930\,nm-980\,nm})$	
Simple ratio water index-1	SRW-1	$R_{858\,nm}/R_{1240\,nm}$	
Simple ratio water index-2	SRW-2	$R_{1070\,nm}/R_{1340\,nm}$	
Moisture stress index	MSI	$R_{1600\,nm}/R_{820\,nm}$	[45]
Moisture stress index-1	MSI-1	$R_{870\,nm}/R_{1350\,nm}$	[44]
Normalized different infrared index	NDI-1	$(R_{850\,nm} - R_{1650\,nm})/(R_{850\,nm} + R_{1650\,nm})$	
Normalized different vegetation index	NDVI	$(R_{858\,nm} - R_{648\,nm})/(R_{858\,nm} + R_{648\,nm})$	
Normalized different water index-1	NDWI	$(R_{858\,nm} - R_{2130\,nm})/(R_{858\,nm} + R_{2130\,nm})$	
Normalized different water index-1	NDWI-Hyp	$(R_{1070\,nm} - R_{1200\,nm})/(R_{1070\,nm} + R_{1200\,nm})$	[46]
Normalized different mass index	NDMI	$(R_{1649\,nm} - R_{1722\,nm})/(R_{1649\,nm} + R_{1722\,nm})$	[47]

R represents spectral reflectance, λ stands for wavelength

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 7 of 24

In Table 2, $R_{\lambda 1}$, $R_{\lambda 2}$, and $R_{\lambda 3}$ represent the spectral reflectance of bands λ_1, λ_2, and λ_3 in the range of 400–2400 nm.

In order to calculate spectral indices in two bands and three bands in batch, we have designed an index calculation software based on the Java environment. The software version is 1.0, and the registration number is 2018Sr281300.

Overall workflow

In this study, the FOD was used for leaf spectral process-

ing, and developed new two-band and three-band indices using the band-optimization algorithm, and collectively referred to multiband spectral indices (MBSIs). Additionally, an integrated index is created by combining
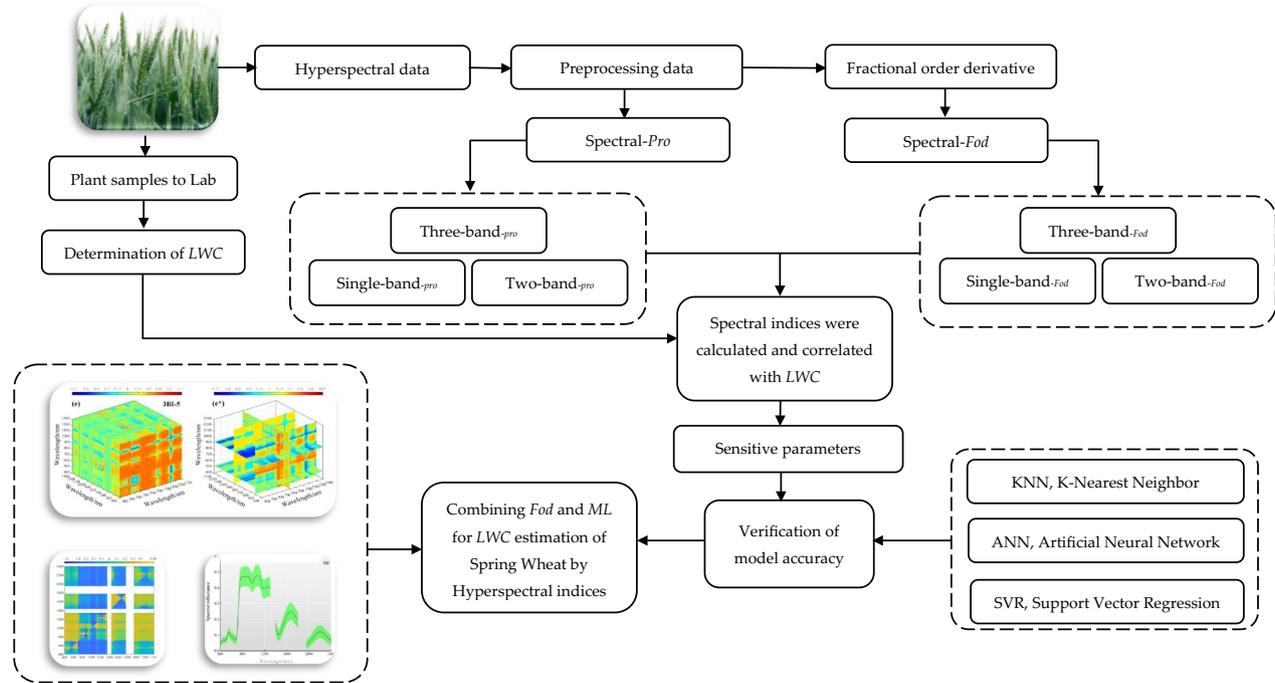
traditional water indices (MIs) with MBSIs to explore whether the combination of different spectral indices improves the accuracy of estimating LWC.

Calibration strategies

The K-nearest neighbor KNN algorithm is used in a variety of classification and regression tasks in machine learning. The key idea behind its machine learning applications is that points tend to share the properties of nearby points [10] (the distance function from one point to another often depends on the context-some

Table 2 The vegetation index of different bands and multiple combinations

Type	Spectral index	Equation	References
Two-band index (2BI)	Ratio vegetation index (RVI)	$R_{\lambda 1}/R_{\lambda 2}$	[54, 55]
	Normalized difference vegetation index (NDVI)	$(R_{\lambda 1} - R_{\lambda 2})/(R_{\lambda 1} + R_{\lambda 2})$	
	Difference vegetation index (DVI)	$R_{\lambda 1} - R_{\lambda 2}$	
Three-band index (3BI)	3BI-1	$R_{\lambda 1}/(R_{\lambda 2} \times R_{\lambda 3})$	[56, 57]
	3BI-2	$R_{\lambda 1}/(R_{\lambda 2} + R_{\lambda 3})$	
	3BI-3	$(R_{\lambda 1} - R_{\lambda 2})/(R_{\lambda 2} + R_{\lambda 3})$	
	3BI-4	$(R_{\lambda 1} - R_{\lambda 2})/(R_{\lambda 2} - R_{\lambda 3})$	
	3BI-5	$(R_{\lambda 2} + R_{\lambda 3})/R_{\lambda 1}$	
	3BI-6	$(R_{\lambda 1} - R_{\lambda 2})/[(R_{\lambda 1} - R_{\lambda 2}) - (R_{\lambda 2} - R_{\lambda 3})]$	
	3BI-7	$(R_{\lambda 1} - R_{\lambda 2}) - (R_{\lambda 2} - R_{\lambda 3})$	

R represents spectral reflectance, λ stands for wavelength

Zununjan *et al. Plant Methods* (2024) 20:97

Page 8 of 24

common ones include Euclidean distance between particles in space, Hamming distance between words, etc.) In a regression setting (where regression is a machine learning technique commonly used to obtain continuous outputs as opposed to discrete outputs in classification), an average (or maximum or minimum) of the KNN is typically used to determine the value of the variable being regressed [76].

Explanation of how KNN works is discussed below: (1) Selecting the optimal value of K; (2) Calculating distance; (3) Finding Nearest Neighbors; (4) Voting for Classification or Taking Average for Regression. In the regression problem, the class label is calculated by taking average of the target values of K nearest neighbors. The calculated average value becomes the predicted output for the target data point.

Artificial neural network (ANN) is a powerful tool used in computer science to solve machine learning problems [58]. It is commonly used for regression and classification tasks. ANN models simulate the electrical activity of the brain and nervous system. In generally, it can fit any non-linear function through a reasonable network structure configuration [59, 77].

ANNs seek to replicate the capabilities of biological neural networks. A node is used to describe an artificial neuron. Like its biologic counterpart, these nodes receive input from synapses and send output when a weight is exceeded [60]. Single-layer ANNs have one layer of input nodes; multilayer ANNs have multiple layers of nodes, including hidden nodes. Both single and multilayer artificial neural networks eventually trigger an output node to fire: this output node makes the decision.

Support vector regression (SVR) is a type of support vector machine (SVM) that is used for regression tasks. It tries to find a function that best predicts the continuous output value for a given input value [61, 62]. The basic steps for building an SVR model are as follows: (1) Data preparation: Collect and preprocess the training data, including feature selection, data cleaning, and normalization; (2) Feature scaling: Scale the input features to ensure they have similar ranges and magnitudes. Common techniques include standardization or normalization; (3) Model selection: Choose an appropriate SVR variant and kernel function; (4) Model training: Use the training data to estimate the model parameters. SVR uses a subset of the training data, called support vectors, to define the regression line or hyperplane; (5) Hyperparameter tuning: Optimize the hyperparameters of the SVR model to improve its performance; (6) Model evaluation: Assess the performance of the trained SVR model using appropriate evaluation metrics.

Verification of model accuracy

Commonly used evaluation indicators for model estimation capability include coefficient of determination (R^2), root mean square errors (RMSE), and relative percent deviation (RPD).

(1) A high coefficient of determination (R^2), indicating a strong linear relationship.
(2) Low Root Mean Square Errors (RMSE) of the model's variables, indicating that the low error between measured and predicted data.
(3) Relative Percent Deviation (RPD), indicating the predictive ability of the model. Its computation process is the ratio between standard deviation (SD) and standard error of prediction (SEP). According to the predictive ability of the model, the RPD is divided into three categories: (I) The value of RPD exceeds 2.0, indicating a model with better predictive ability. (II) The RPD values ranging from 1.4 to 2.0 represent a model with general predictive ability. (III) The RPD value is less than 1.4, indicating that it has poor predictive ability.

A model with an R^2 approaching 1, an RMSE approaching 0, and an RPD greater than or equal to 2.0 exhibits estimation ability and stability [49].

$$R^2 = \left[\frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 + \sum_{i=1}^{N}(y_i - \overline{y})^2}} \right]^2,$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\gamma_i - \beta_i)^2}{n}},$$

$$RPD = SD/SEP.$$

Note: x_i and y_i are measured and predicted values, respectively; $\overline{x}$ and $\overline{y}$ represent the means measured and predicted values, respectively; and n is the number of samples. SD and SEP represent the standard deviation and Standard error of prediction, respectively.

Results

Leaf water content and FOD hyperspectral curves

Figure 4 displays the statistical characteristics of the measured LWC for the entire dataset, including the calibration and validation subsets. The analysis reveals that the whole dataset exhibited a wide variation, with a minimum, maximum, and coefficient of variation (CV) of 64.94%, 91.54%, and 4.58%, respectively, indicating a diverse range of soil samples in the study area. The range of LWC contents observed in the validation dataset was
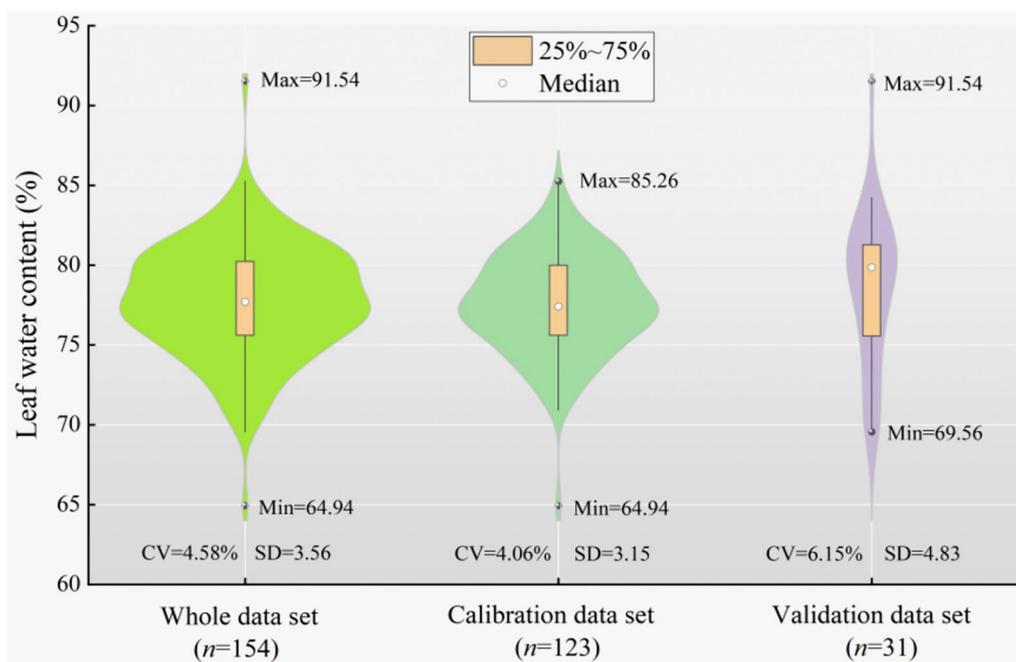
Fig. 4 Descriptive statistics of spring wheat leaf water content. Max, Min, CV and SD represent the maximum, minimum, coefficient of variation and standard deviation, respectively

consistent with that of the calibration dataset. The mean, standard deviation (SD), and CV values from these three datasets were relatively similar, suggesting that the calibration and validation subsets are representative of the entire population.

The spectral profiles of the spring wheat canopy corresponding to different moisture contents of leaves are presented in Fig. 5a. The reflectance changed due to the difference in leaf water content, and a decreasing trend was observed in spectral reflectance with an increase in leaf water content. The slopes at 930–970 nm and

1100–1200 nm increased with an increase in leaf water content. Among them, wavelengths of approximately 970 nm and 1200 nm were considered better choices for estimating leaf water content due to the absence of atmospheric disturbance.

Figure 5b shows the canopy spectral profiles of spring wheat treated with differentiation at order 0–2. Each order of differential spectral profiles had a gradual process with an increase in order, resulting in a decrease in spectral reflectance. The commonly used first- and second-order differentiation of spectral profiles differs
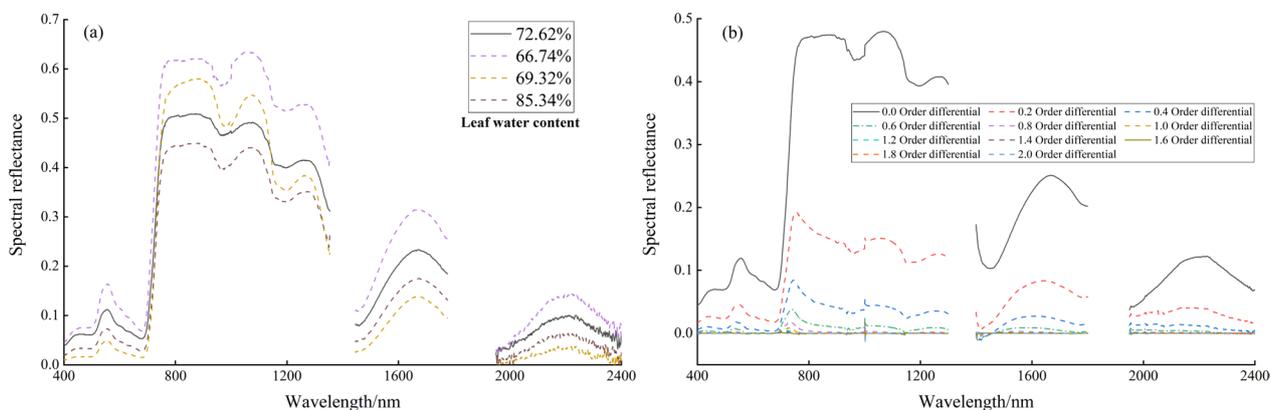
Fig. 5 Canopy spectral curves of spring wheat. **a** Canopy spectral curves of spring wheat based on different water contents. **b** Canopy spectra of spring wheat with 0.0–2.0 order differentials

significantly from the original one, and the information between them may be missed. On the other hand, fractional differentiation can exploit intermediate information, which can further extract and utilize hyperspectral information.

One-dimensional correlation analysis based on FOD

Based on the above analysis, a Pearson correlation analysis was conducted on the water content of spring wheat leaves and 0–2 order differential reflectance, resulting in the distribution of correlation coefficients at each wavelength. As illustrated in Fig. 6a, the differential spectral curves of each order displayed a gradual change with increasing differential order, whereas the correlation coefficient curves exhibited increasing fluctuations and lacked strong regularity. The number of bands whose correlation coefficients passed the 0.01 level significance test displayed an initial increasing and then decreasing trend.

The 0–0.8 order differential spectral curves displayed little fluctuation and similar shapes. The number of bands with correlation coefficients gradually increased by passing the 0.01 level significance test (critical value ± 0.201), mainly concentrated in the range of 700–1100 nm, and exhibited positive and negative correlations with leaf water content (Fig. 6a). The maximum correlation coefficient appeared at 778 nm (Fig. 6b) using a 0.8-order differential, reaching 0.35. The fluctuation of 1.0–2.0 order differential spectral curves gradually increased, and the number of bands with correlation coefficients passing the 0.01 horizontal significance test gradually decreased (Fig. 6c), with the absolute values of correlation coefficients decreasing gradually. The absolute values of the correlation coefficients displayed an initial increase and then a decreasing trend, with the peak value appearing in the 0.8-order differential spectral reflectance. This finding indicates that the best order between 400 and 2400 nm is the 0.8 order.

The above results show that the fractional differential of the spring wheat canopy spectrum has some advantages in screening sensitive bands compared with the commonly used first- and second-order differentials and can more accurately find the bands that have a higher correlation with LWC.

Two-dimensional correlation analysis based on FOD

The correlation between 12 traditional moisture indices and LWC is shown in Table 3. The analysis showed that the water vegetation index and LWC had a significant correlation, both passing the 0.01 significance level ($\rho < 0.01$), and the correlation coefficient is between
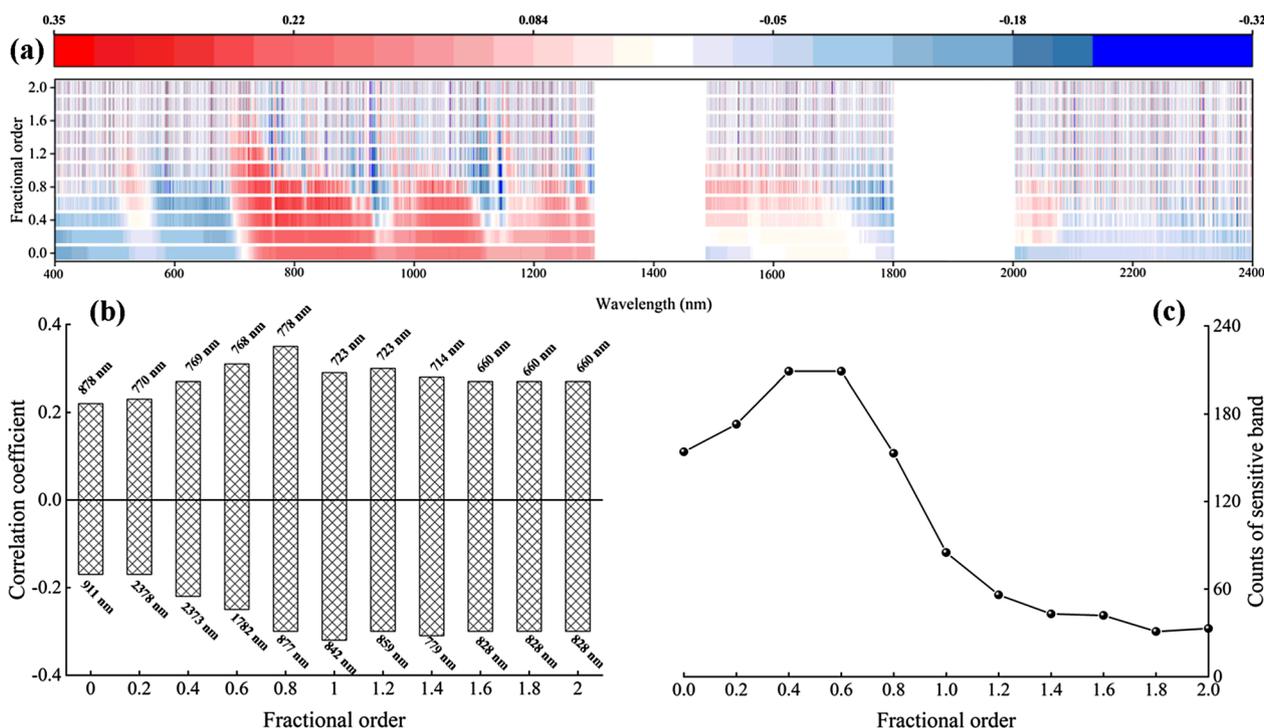
Fig. 6 Color map of correlation coefficients (*r*) and statistical analysis. **a** Correlation coefficients between LWC and spectral reflectance based on 0.0 to 2.0 order differentials. **b** Maximum-minimum correlation coefficient for each order. **c** Number of bands with correlation coefficients tested by 0.01 level significance

Zununjan *et al. Plant Methods*　　(2024) 20:97

Page 11 of 24

Table 3 Correlation coefficient between moisture indices and LWC

Variable	Correlation coefficient	Variable	Correlation coefficient
WI	+0.22**	MSI-1	0.24**
WBI	−0.22**	NDI	0.20**
FWBI	−0.33**	NDVI	0.23**
SRWI-1	+0.18*	NDWI	0.19*
SRWI-2	+0.25**	NDWI-Hyp	0.21**
MSI	−0.20**	NDMI	0.13

** and * represent the 0.01 level and 0.05 level, respectively

−0.22 and 0.33. The correlation coefficient between the FWBI and LWC is 0.33**, which means that the correlation between the FWBI and LWC is higher than that of the traditional water spectral index. The results show that the performance of the traditional water spectral index

varies with geographic environment. Therefore, it is necessary to find a better spectral vegetation index for sensitive combinations.

Pearson correlation analysis was conducted between the DVI and LWC utilizing fractional differential processing, and the results are presented in the form of a heatmap in Fig. 7. An analysis of Fig. 7I. revealed that the DVI calculated at each order displayed varying degrees of sensitivity to LWC. Specifically, the correlation coefficient of the DVI to LWC calculated at orders 0–0.8 first decreased and then increased, reaching its highest correlation at order 0.8 with an absolute correlation coefficient value of 0.69. For calculations at orders 1–2, the correlation coefficient of DVI on LWC gradually decreased, and the absolute values of the correlation coefficient ranged from 0.46 to 0.69.

An analysis of Figs. 8I and 9I revealed that the NDVI and RVI, respectively, calculated at each order exhibited varying degrees of sensitivity to LWC. Specifically,
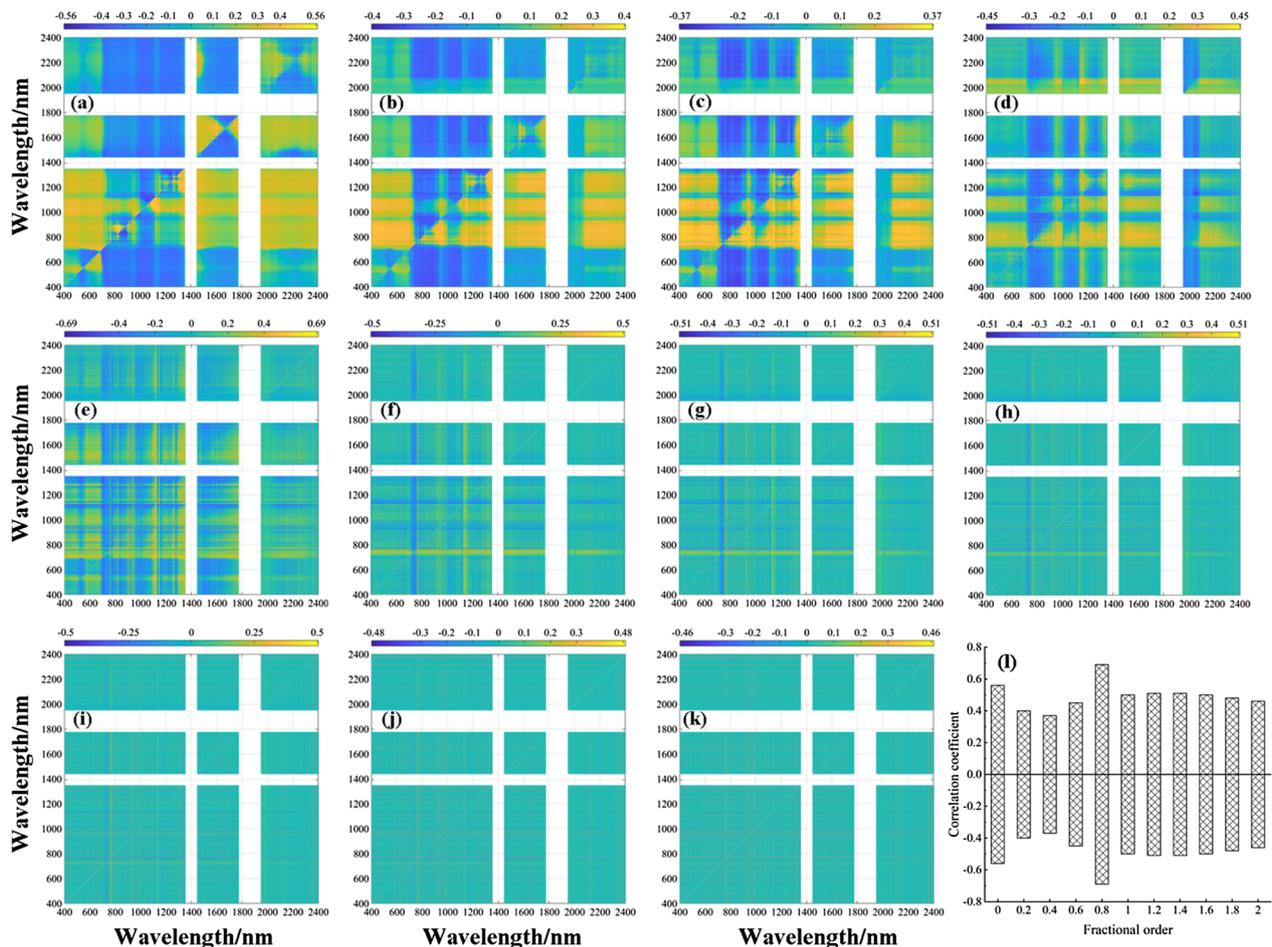
Fig. 7 Two-dimensional correlation coefficients (*r*) between LWC and DVI in the training dataset at eleven derivative orders. **a** Raw data, **b** 0.2-order, **c** 0.4-order, **d** 0.6-order, **e** 0.8-order, **f** 1.0-order, **g** 1.2-order, **h** 1.4-order, **i** 1.6-order, **j** 1.8-order, **k** 2.0-order, and **l** maximum–minimum correlation coefficient for each order

the NDVI and RVI calculated at orders 0–0.8 showed a decreasing trend followed by an increasing trend in the correlation coefficient of LWC and displayed the highest correlation at order 0.8, with absolute correlation coefficients of 0.68 and 0.71, respectively. For calculations at orders 1.0–2.0, the correlation coefficients of the NDVI and RVI for LWC gradually decreased, with absolute correlation coefficient values ranging from 0.32 to 0.68 and 0.34 to 0.71, respectively. The above analysis indicates that the RVI, NDVI, and DVI calculated using 0 to 0.8-order spectral reflectance presented higher correlation coefficient in Figs. 7, 8I, 9. The band combination tested for significance at 0.01 mainly focused on the range of 400–1300 nm, suggesting that this band region can provide more band combination information for changes in leaf moisture. Therefore, the next step is to utilize the 400–1300 nm band information to calculate the

three-band combination vegetation index and examine its sensitivity trend to LWC.

Three-dimensional correlation analysis based on FOD

The correlation analysis between fractional-order differential spectroscopy and the spectral index revealed that spectral reflectance based on 0.8-order pretreatment exhibited the highest correlation with LWC. The sensitive range of the two-band spectral index based on the 0.8-order differential to the LWC was 400–1300 nm. To calculate the three-band spectral index, 0.8-order spectral data within the range of 400–1300 nm were selected. Figure 10**.** shows the correlation heatmap between the three-band spectral index and LWC. Specifically, Fig. 10a–g represent the correlation results between LWC and all possible combinations of three-band vegetation indices within the range of 400–1300 nm, while Fig. 10a*–g* represent the best
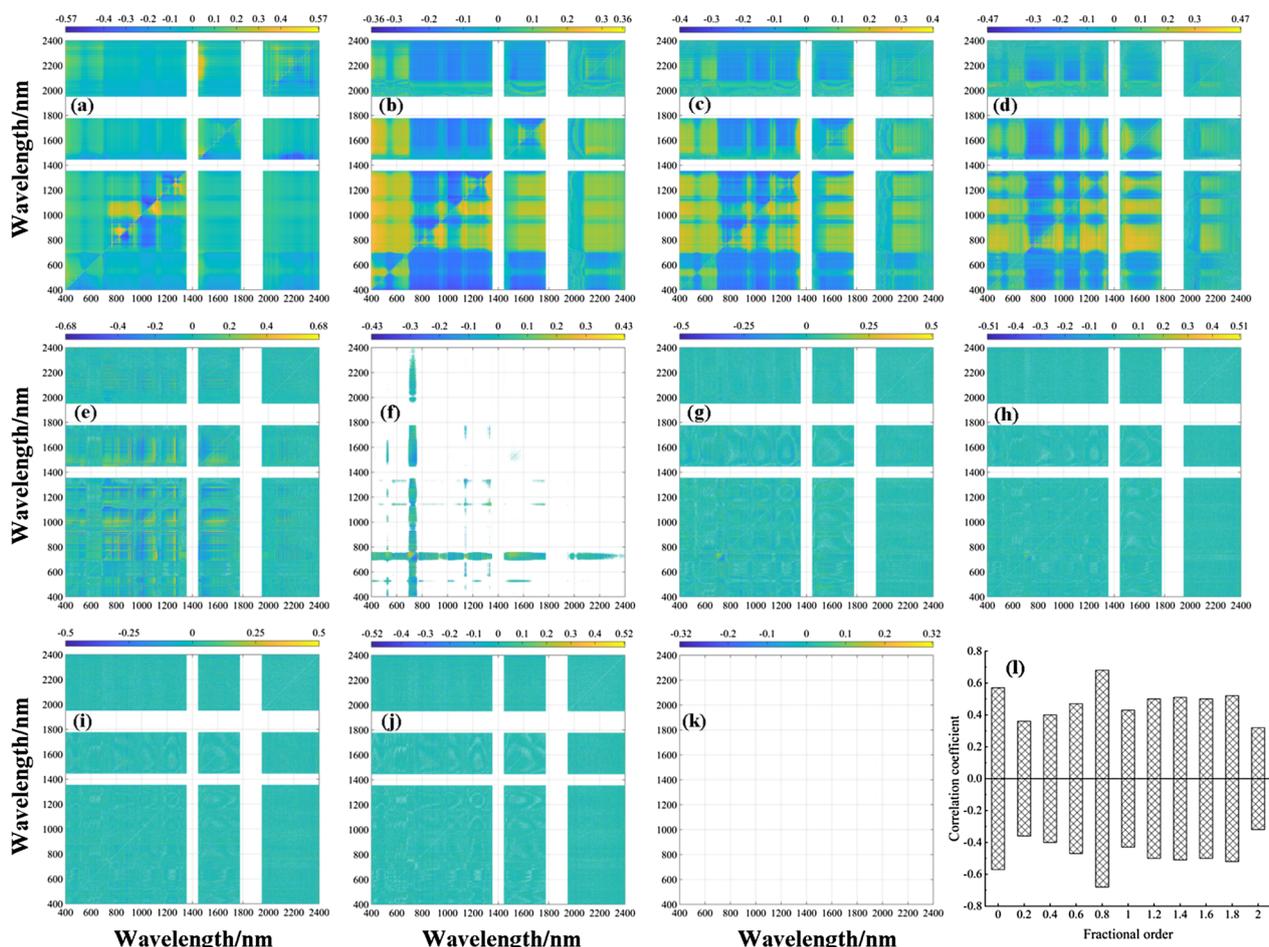
Fig. 8 Two-dimensional correlation coefficients (*r*) between the LWC and NDVI in the training dataset at eleven derivative orders. **a** raw data, **b** 0.2-order, **c** 0.4-order, **d** 0.6-order, **e** 0.8-order, **f** 1.0-order, **g** 1.2-order, **h** 1.4-order, **i** 1.6-order, **j** 1.8-order, **k** 2.0-order, and **l** maximum–minimum correlation coefficient for each order. (Note: the reason for the white areas in Fig. 11-k is due to limited correlation data and a dispersed distribution, making it difficult to discern the presence of data in the mapping results)
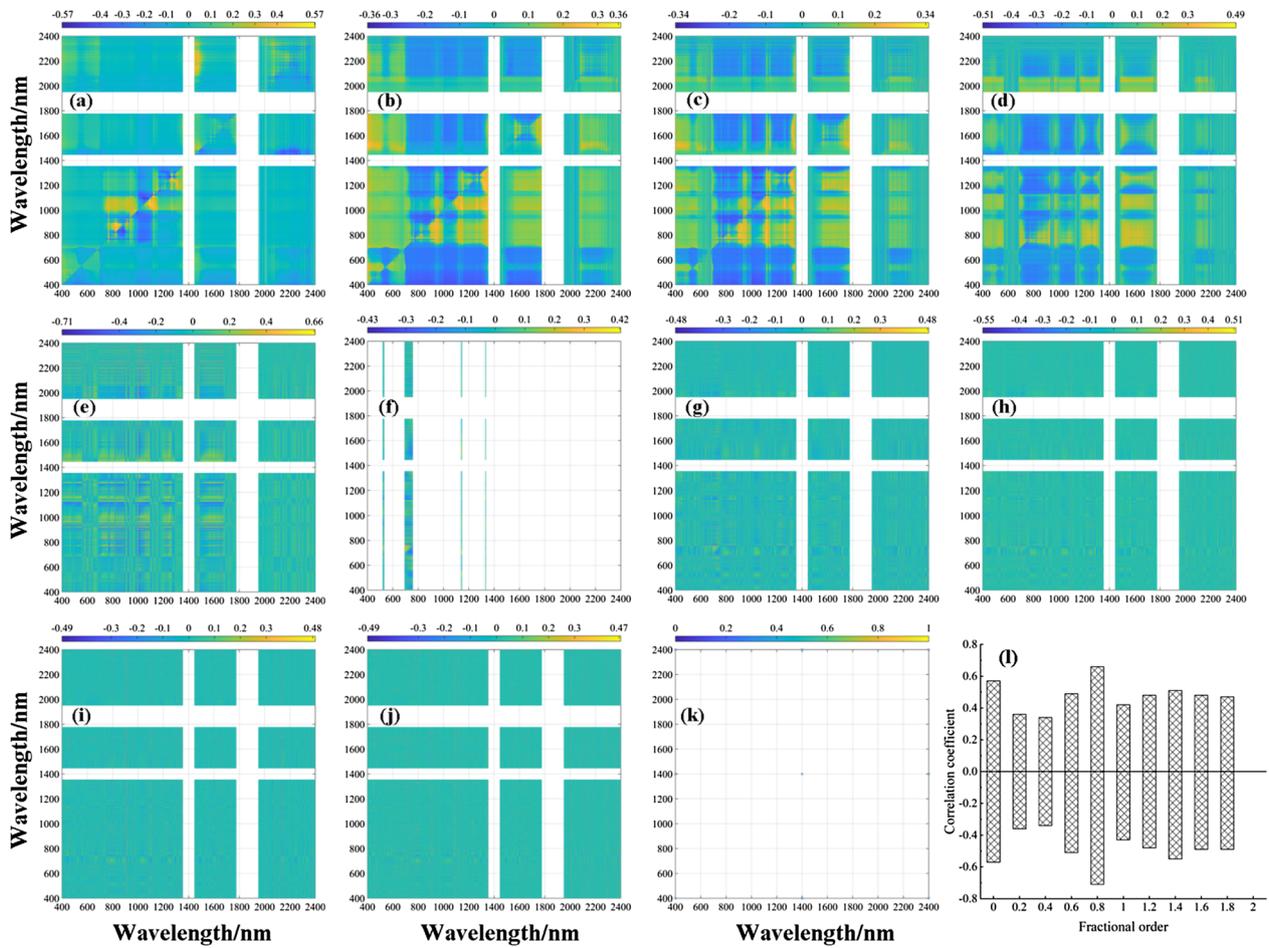
Fig. 9 Two-dimensional correlation coefficients (*r*) between LWC and RVI in the training dataset at eleven derivative orders. **a** Raw data, **b** 0.2-order, **c** 0.4-order, **d** 0.6-order, **e** 0.8-order, **f** 1.0-order, **g** 1.2-order, **h** 1.4-order, **i** 1.6-order, **j** 1.8-order, **k** 2.0-order, and **l** maximum–minimum correlation coefficient for each order (the reason for the white areas in Fig. 12k is due to limited correlation data and a dispersed distribution, making it difficult to discern the presence of data in the mapping results)

three-band vegetation index, which is the combination of the most sensitive bands to LWC correlation.

After conducting a sensitivity analysis on three band vegetation indices with respect to LWC, the associations between 3BI-1 and 3BI-2 with LWC were ± 0.51 and ± 0.54, respectively. However, there was no significant increase in the correlation coefficient when compared to vegetation indices in the two bands. On the other hand, 3BI-3, 3BI-4, 3BI-5, 3BI-6, and 3BI-7 showed significantly higher correlations with LWC, all of which had values above 0.70. Among them, 3BI-7 exhibited the highest correlation with LWC, with an absolute coefficient value of 0.76. Overall, the correlation coefficient showed some improvement when compared to vegetation indices in the two bands.

Combining the data from Fig. 10, the optimal three band combination information was statistically determined, and among the operations of spectral reflectance

based on the 0.8-order differential treatment for three band vegetation indices, the optimal three band combination information is shown in Table 4. The reflectance corresponding to the wavelengths (929 nm, 851 nm, and 446 nm) was calculated in the form of the $(R_{929\ nm} - R_{851\ nm}) - (R_{851\ nm} - R_{446\ nm})$ combination, which could enhance the susceptibility degree to LWC, indicating that the single band information was weaker than the result of the combined band information and that the three band vegetation index was better than the two band vegetation index in terms of the combined band information.

LWC prediction model based on ML

According to the correlation coefficient values and significance test ($p < 0.01$), the important wavebands were identified and extracted efficiently. K-nearest neighbor (KNN), Artificial neural network (ANN) and support vector machine (SVM) models were constructed to
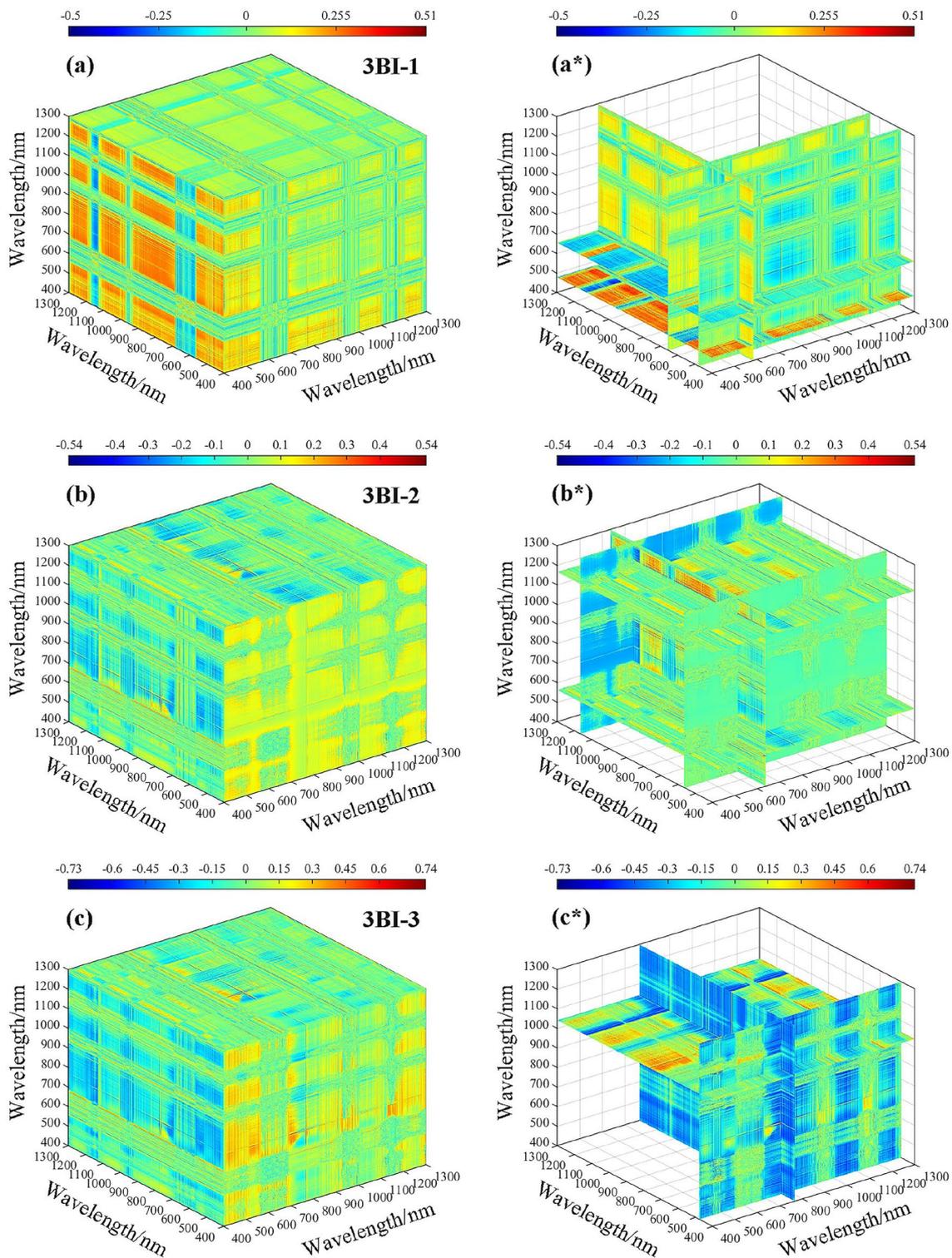
Fig. 10 Three-dimensional correlation coefficients (*r*) between the LWC and three-band indices based on 0.8 order differentials. **a** 3BI-1, **b** 3BI-2, **c** 3BI-3, **d** 3BI-4, **e** 3BI-5, **f** 3BI-6, and **g** 3BI-7
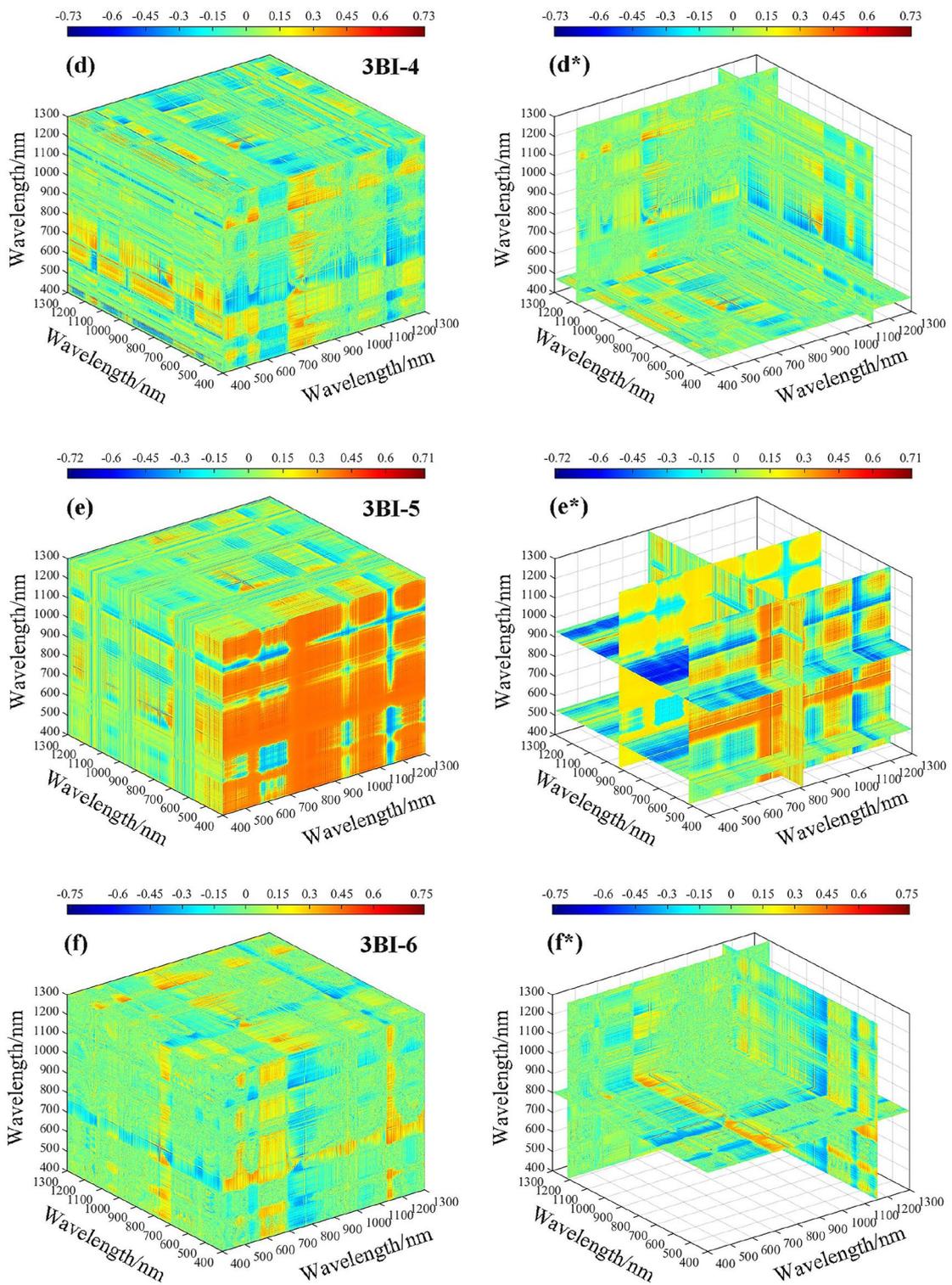
Fig. 10 continued
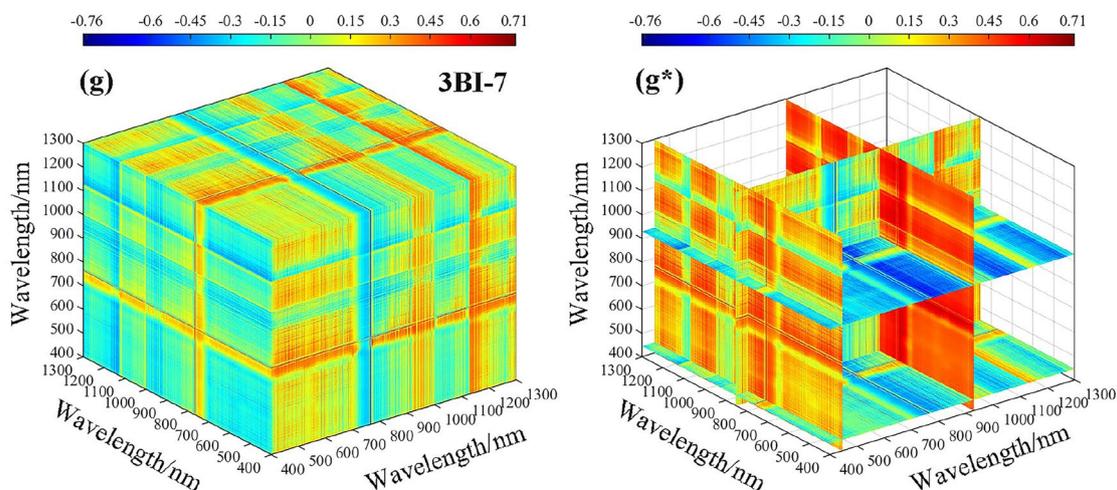
Fig. 10 continued

Table 4 Optimum three-band combination information statistics

Type	Fractional order	Spectral index	Optimum combination band	r
Three-band index (3BI)	0.8-order	3BI-1	583 nm 656 nm 479 nm	0.51
		3BI-2	636 nm 1167 nm 564 nm	0.54
		3BI-3	766 nm 478 nm 1042 nm	0.74
		3BI-4	1129 nm 1175 nm 471 nm	0.73
		3BI-5	814 nm 929 nm 525 nm	-0.72
		3BI-6	1156 nm 1214 nm 802 nm	0.75
		3BI-7	929 nm 851 nm 446 nm	-0.76

"*r*" represents the correlation coefficient

quantify the LWC, and an independent validation dataset was used to verify the quantitative capabilities of the models. In this study, 27 models were constructed based on machine learning, and the performances of all models are shown as scatter plots (Fig. 11).

Research shows that the model accuracy R^2_{pre} of the sensitive single-band reflectance estimation model of LWC can reach 0.32, and the RMSE$_{pre}$ is 5.03%. The model accuracy R^2_{pre} of the spectral data after fractional order differential processing can reach 0.61, and the RMSE$_{pre}$ is 2.49%. The model accuracy R^2_{pre} of estimating LWC using the 12 existing spectral indices reached 0.37, and the RMSE$_{pre}$ was 5.43%. Compared with the single-band reflectance, the model accuracy was improved. The model accuracy R^2_{pre} of estimating with two-band optimized spectral indices (DVI, RVI and NDVI) was the highest, reaching 0.64, and the RMSE$_{pre}$ was 2.35% (Table 5). The accuracy of the estimation model was significantly enhanced, indicating that optimizing the combination of spectral indices can improve the model's estimation ability.

After 0.8 spectral processing, the model accuracy R^2_{pre} of the constructed three-band spectral index estimation reached 0.83, and the RMSE$_{pre}$ was 2.21%. The model accuracies R^2_{pre} of FWBI+2BI-$_{FOD}$ and FWBI+3BI-$_{0.8}$ reached 0.81 and 0.86, respectively, and the RMSE$_{pre}$ values were 2.25% and 2.11%, respectively. The existing spectral index and the MBSI index based on fractional order differentiation significantly improved the estimation accuracy and reduced the error.

The results of the accuracy verification of the estimated LWC model constructed via sensitive bands, the existing moisture indices, two-band and three-band optimized indices. Screening was conducted on 27 estimation models constructed for 9 data types, and the results of the best estimation models corresponding to each data type were added to the part. Scatter plots of measured and predicted values were showed in Fig. 11. Among the 27 models, the models based KNN and FOD algorithm showed better prediction ability. In this study, the FWBI-3BI-0.8 order based on KNN illustrated the highest R^2 accuracy, the lowest error (RMSE) and the greatest estimation ability (RPD).
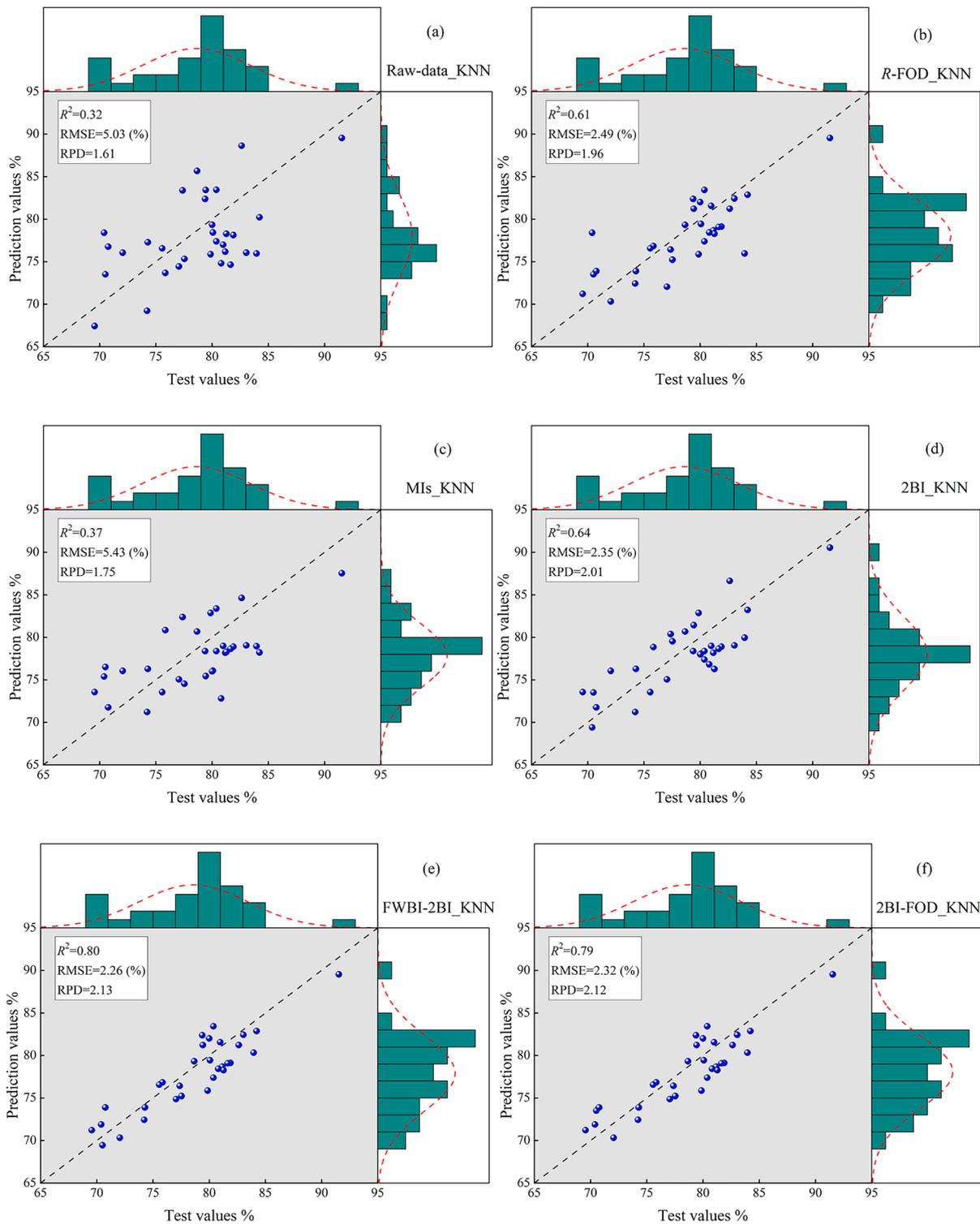
Fig. 11 Scatter plot of measured and predicted values. **a** The KNN estimation model based on Raw-data, the wavelength (769 nm–924 nm). **b** The KNN estimation model based on R_{-FOD}, the wavelength (711 nm–8754 nm).the FWBI and 2BI$_{-FOD}$. **c** The KNN estimation model based on exist moisture indices. **d** The KNN estimation model using two-band optimized indices. **e** The KNN estimation model using moisture index and two-band optimized index. **f** The KNN estimation model using two-band optimized indices based on FOD. **g** The KNN estimation model using moisture index and two-band optimized indices based on FOD. **h** The KNN estimation model using three-band optimized indices based on 0.8-order. **i** The KNN estimation model based on the FWBI and three-band optimized indices based on 0.8-order

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 18 of 24
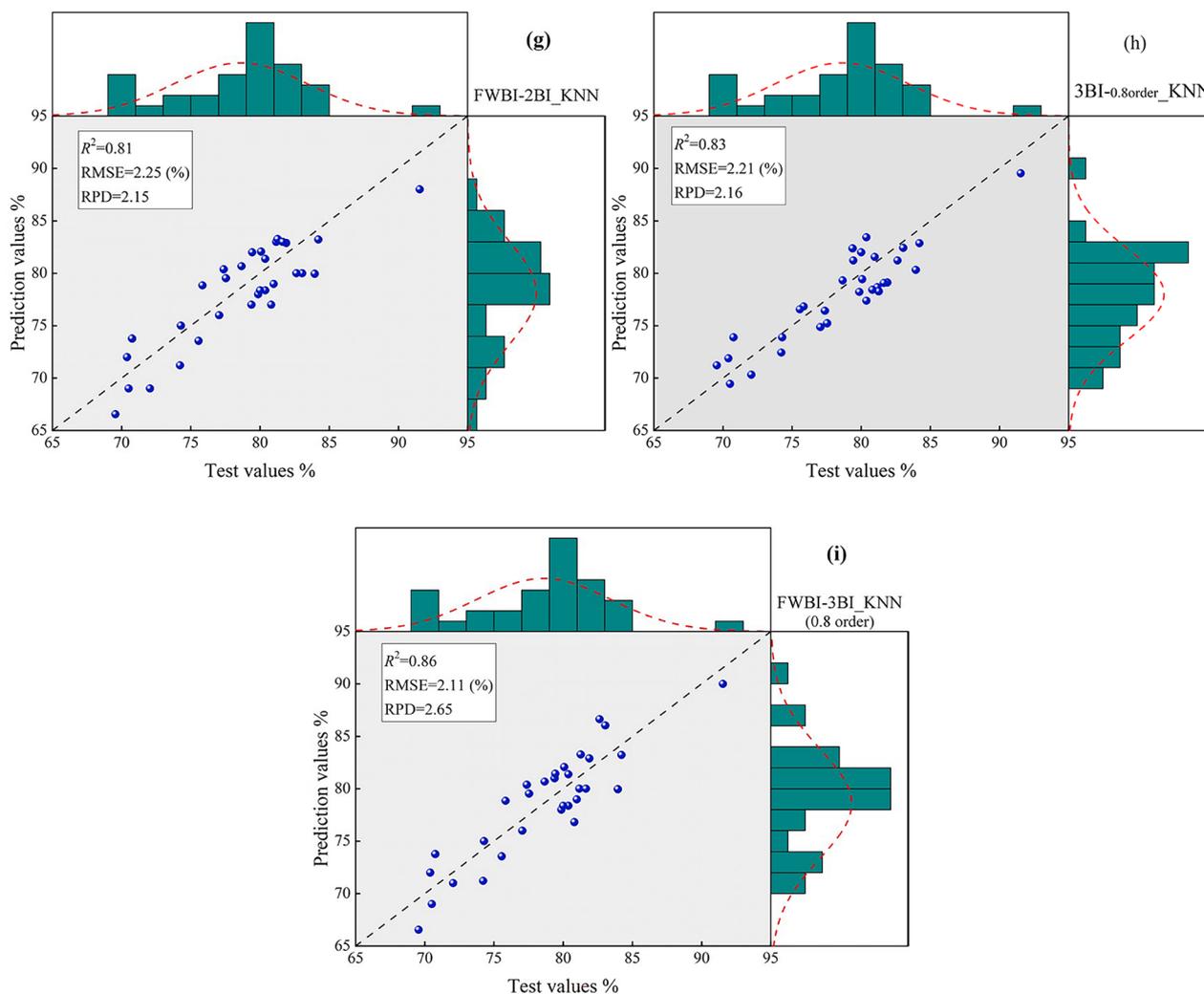
Fig. 11 continued

Discussion

Hyperspectral fulfilled the characteristics of multiple bands, strong continuity, and large information [64]. However, it was easy to produce overfitting problems, which can affect the predictive performance of models [65]. Therefore, in the Vis–NIR analysis of spring wheat properties, it is important to seek efficient methods to process the raw spectrum and reduce the number of redundant bands has become an important focus [65, 66].

Spectral derivative preprocessing techniques can remove baseline drift effects, reduce overlapping spectral bands, solve overlapping peaks, improve spectral resolutions and sensitivities, and eliminate interferences resulted from other background factors [31]. In previous studies, integer order derivative was commonly used for preprocessing the raw spectrum [49, 64, 65]. Wherein FOD could vary at a small interval to keep the spectral information changing slowly, further extract the effective information and allow us to detect more characteristics of certain spectral signal than integer order derivatives [18, 49]. And then some useful spectral information should not be ignored. The use of FOD improved correlation between LWC and spectrum, the best correlation coefficient for LWC was achieved using 0.8-order reflectance (Fig. 6b). These results provide the potential to establish a more LWC estimation model.

Optimal spectral indices, which are calculated based on the sensitive wavebands related to characteristic attributes, can easily detect subtle absorption peaks and can be used to predict different spring wheat properties [63]. A number of algorithms have been proposed to optimize two-band and three-band combinations to obtain spring

Table 5 The accuracy and validation of the models based on multiband spectral data

Data-type	Variable	Method	Training dataset		Validation dataset		RPD
			R^2_{sim}	RMSE$_{sim}$	R^2_{pre}	RMSE$_{pre}$	
Raw-data	The reflectance of 154 corresponding to the wavelength (769–924 nm)	KNN	0.41	4.06	0.32	5.03	1.61
		SVR	0.25	5.21	0.21	5.62	1.50
		ANN	0.19	5.39	0.11	6.07	1.33
R_{-FOD}	Reflectance of 165 at 0.8 order for wavelength 711–875 nm	KNN	0.65	2.42	0.61	2.49	1.96
		SVR	0.55	3.11	0.48	4.01	1.92
		ANN	0.46	4.18	0.37	4.56	1.84
MIs	The existing 12 moisture indices	KNN	0.46	4.12	0.37	5.43	1.75
		SVR	0.21	5.19	0.17	5.71	1.69
		ANN	0.22	4.68	0.16	6.26	1.55
2BI	DVI$_{(1046, 1057 nm)}$ RVI$_{(1272, 1279 nm)}$ NDVI$_{(1272, 1279 nm)}$	KNN	0.67	2.56	0.64	2.35	2.01
		SVR	0.45	3.47	0.40	4.23	1.88
		ANN	0.39	4.21	0.28	4.86	1.76
FWBI-2BI	FWBI$_{(900,930,980 nm)}$ RVI$_{(1272, 1279 nm)}$	KNN	0.82	2.21	0.80	2.26	2.13
		SVR	0.78	2.28	0.74	2.32	2.09
		ANN	0.81	2.24	0.79	2.27	2.11
2BI$_{FOD}$	DVI$_{(698, 1274 nm)}$ RVI$_{(1156, 1628 nm)}$ NDVI$_{(1182, 1149 nm)}$	KNN	0.81	2.26	0.79	2.32	2.12
		SVR	0.79	2.31	0.72	2.38	2.06
		ANN	0.81	2.29	0.78	2.34	2.08
FWBI-2BI$_{FOD}$	FWBI$_{(900,930,980 nm)}$ RVI$_{(1156, 1628 nm)}$	KNN	0.84	2.20	0.81	2.25	2.15
		SVR	0.80	2.28	0.76	2.36	2.08
		ANN	0.82	2.24	0.79	2.30	2.11
3BI$_{-0.8 order}$	3BI-1$_{(583,656,479 nm)}$ 3BI-2$_{(636,1167,564 nm)}$ 3BI-3$_{(766,478,1042 nm)}$ 3BI-4$_{(1129,1175,471 nm)}$ 3BI-5$_{(814,929,525 nm)}$ 3BI-6$_{(1156,1214,802 nm)}$ 3BI-7$_{(929,851,446 nm)}$	KNN	0.85	2.18	0.83	2.21	2.16
		SVR	0.81	2.23	0.79	2.35	2.11
		ANN	0.83	2.20	0.80	2.24	2.13
FWBI-3BI$_{-0.8 order}$	FWBI$_{(900,930,980 nm)}$ 3BI-7$_{(929,851,446 nm)}$	KNN	0.89	2.05	0.86	2.11	2.65
		SVR	0.85	2.24	0.83	2.26	2.13
		ANN	0.86	2.19	0.84	2.21	2.19

"*sim*" stands for simulation, "*pre*" stands for prediction

wheat properties of interest. In this study, we optimized the two-band and three band spectral indices based on FOD. According to the correlation analysis, the LWC data and optimized indices (two/three-band indices) have a better correlation, correlation coefficients (*r*) of RVI$_{(1156 nm, 1628 nm)}$ was $-0.71**$ and 3BI-7$_{(929 nm, 851 nm, 446 nm)}$ was $-0.76**$, respectively, and the optimized indices based on FOD was significantly better than that of the exist indices (shown in Fig. 12). The results showed that there was great potential in using the band optimization algorithm and FOD to estimate LWC.

In this study, the purpose of calculating two-dimensional and three-dimensional indices is to effectively integrate relevant spectral signals and identify better band combinations. The two-band and three-band indices encompass the visible and near-infrared spectral ranges. It might be that moisture content can alter the

absorption and reflection properties of leaves in the visible spectrum, while in the near-infrared spectrum, moisture content primarily affects light scattering. As leaf moisture is lost, the biological activities of the leaves are inhibited, preventing normal leaf functioning, leading to an increase in reflectance in the spectral curve of leaves in the visible and near-infrared regions as moisture loss progresses [73].

Simultaneously, we constructed 27 models for estimating the LWC in spring wheat based on multiband spectral data (single bands, moisture indices, two-band spectral indices and three-band spectral indices) and machine learning (KNN, ANN and SVR). And chosen 3 metrics to evaluate the performance of the algorithm, including R^2, RMSE, and RPD. The range of R^2 is [0, 1], where $R^2 = 1$ indicates the model perfectly predicted the data, and $R^2 = 0$ indicates the model cannot explain the
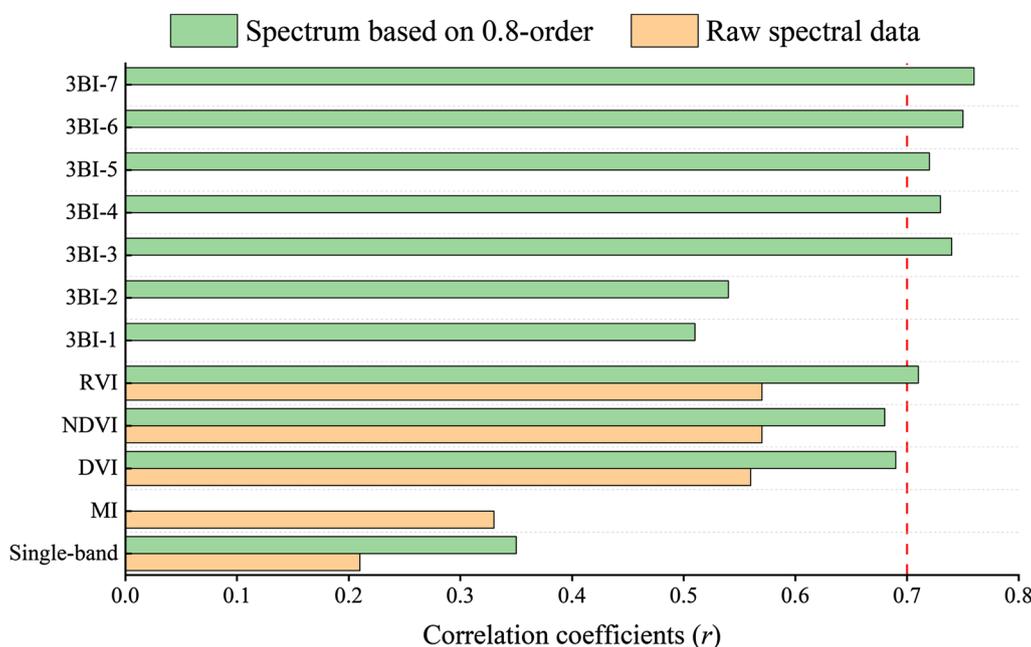
Fig. 12 The relationship between multiband spectral data and LWC

variance. RMSE is a metric used to measure the predictive accuracy of a predictive model on continuous data. The value of RPD exceeds 2.0, indicating a model with better predictive ability.

The integrated index (MBSI) allows high-precision estimation of LWC, which acquired the highest coefficient of determination, the lowest root mean square error and the best predictive ability ($R^2 = 0.86$, RMSE = 2.11% and RPD = 2.65). There is some variability in the fitting accuracy of all models, and the estimation capability of models could be ranked as follows: $Model_{(KNN\ based\ FWBI+3BI+0.8-order)} > Model_{(KNN\ based\ 3BI+0.8-order)} > Model_{(KNN\ based\ FWBI+2BI+FOD)} > Model_{(KNN\ based\ 2BI+FOD)} > Model_{(KNN\ based\ FWBI+2BI)} > Model_{(KNN\ based\ 2BI)} > Model_{(KNN\ based\ R-FOD)} > Model_{(KNN\ based\ MIs)} > Model_{(KNN\ based\ Raw\ data)}$.

The reasons may be described as follows:

(1) Spectral information is often easily affected by the leaf surface, leaf structure (lignin, cellulose, etc.) and the external environment, and the raw spectral data contains complex interference information. Simultaneously, the spectral information has a certain regional nature due to the variety and type of plants. Therefore, the estimation accuracy of single-band reflectance and the existing 12 moisture indices is different.

(2) The spectral data can be refined by fractional differential processing and highlight the hidden information, which is helpful for screening the most sensitive bands, and therefore, the two-band and three-band optimized spectral indices can highlight the most influential indicators in the region, thus eliminating noise information and obtaining a more effective models.

Overall, our results show that there is great potential in using the FOD and ML to estimate LWC, and the optimal model accuracy was comparable to those reported from studies. Alireza Sharifi et al. used the Sentinel-1 SAR data and three methods (MLR, RVR and SVR) to estimate rice parameters. The results indicated that the nonparametric methods (SVR and RVR) is much better than that of the parametric regression (MLR) for rice parameter estimations [78]. Alireza Sharifi et al. used Transformed chlorophyll absorption in reflectance index (TCARI) and Modified chlorophyll absorption in reflectance index (MCARI) to determine crop nutrition status. The results indicated that the performance of TCARI and MCARI was allowed the creation of high accuracy crop nutrition maps, the use of the near infra-red and red-edge bands led to better results [79]. The findings of this study are similar in comparison, as the band combinations in this study also focus on the near-infrared and shortwave infrared regions, highlighting the performance ability of spectral indices. Previous research on the estimation of the moisture content of crop leaves from visible and near-infrared reflectance was shown in Table 6.

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 21 of 24

After analyzing the methods of constructing models and predicting accuracy, four models, Model-$_{WC+FOD+ANN}$, Model-$_{RVI437,466\ nm+NDVI747,1956\ nm+BPNN}$, Model-$_{x-LW+KNN}$ and the integrated Model-$_{MBSI+MI+FOD+KNN}$, were found to have higher estimation capabilities. Except for the comparative Model-$_{RVI437,466\ nm+NDVI747,1956\ nm+BPNN}$, Model-$_{MBSI+MI+FOD+KNN}$ has higher estimation accuracy. Thus, the best performance of Model-$_{MBSI+MI+FOD+KNN}$ can realize the regional-scale monitoring of wheat canopy water status. Xuenan Zhang et al. used a machine algorithm of gradient boosted decision tree (GBDT) based on the combination of $ND_{(1287,1673)}$ and crop water stress index (CWSI), the optimal prediction accuracy ($R^2 = 0.86$, RMSE $= 0.01$) of rice LWC was produced. In previous studies, few researchers have considered extending the spectral index method to more than two bands, particularly optimizing three-band indices and estimating crop leaf moisture by combining traditional spectral indices.

Currently, domestic and foreign research has achieved certain results, demonstrating the feasibility of using hyperspectral technology to monitor crops in farmlands. Previous studies often used complex algorithms to construct estimation models and screen feature bands. For instant, the precision of estimating LWC in winter wheat by combining stepwise regression method and partial least squares (SRM-PLS) or PLS based on the relational degree of grey relational analysis (GRA) between water vegetation indexes (WVIs) and LWC [67, 68]. However, such algorithms have poor operability and are not conducive to obtaining feature bands intuitively and efficiently. In this study, the computational complexity of our proposed algorithm are slightly more easy; we extracted moisture-sensitive bands by preprocessing the raw spectrum using spectral information from different leaf positions and canopy layers to select the optimal spectral characterization information for wheat moisture status. We used the optimal modeling method for feature band selection to construct the optimal wheat moisture estimation models. Moreover, we achieved good results in the sensitivity band screening of the wheat leaf moisture spectral estimation model, the influence of the number of modeling bands on the accuracy of the estimation model, and the monitoring of moisture spectra in different leaf positions. This result indicates that in future research, it will be necessary to develop and utilize spectral pretreatment techniques to reduce the spectral response of LWC in order to achieve the purpose of rapid and nondestructive estimation of other spring wheat property parameters. Esmaeili et al. proposed the band selection method based on CNN embedded GA CNNeGA, the evaluation of the proposed method and the obtained results are satisfactory [81]. This method also provides strong assistance for our future screening of hyperspectral data and spectral indices.

Limitations of this study: Spring wheat exhibits different hyperspectral characteristics in different regions due to the influence of climate, leaf size, shape, growth period and other factors, therefore, many factors must be considered when selecting the leaf moisture sensitivity index of spring wheat. First, the data collected from 154 sampling points is limited, and further increasing the data volume is necessary to ensure the stability and applicability of the estimation model. Second, this study is limited to the heading stage of wheat, and it is necessary to validate the three-band indices and models in each growth stage, lacking systematic data research and modeling. Addressing these limitations will be a key focus of future research in this study.

Table 6 Comparison of the estimation of LWC in wheat using hyperspectral data

Study area	Year	Methods	R^2	References
Changping Beijing, China	2021	WI-4 + NDWSI-4 + Linear regression	0.82	[8]
Jiaozuo City, Henan, China	2023	WC + FOD + ANN	0.86	[9]
Yangzhou, China	2021	$RVI_{437,466\ nm}$ + $NDVI_{747,1956\ nm}$ + BPNN	0.88	[2]
Luoyang Henan, China	2021	x-LW + KNN	0.84	[11]
Fukang Xinjiang, China	2019	FD + GRA + PLSR	0.81	[13]
Astaneh-ye Ashrafiyeh region in the north of Iran	2019	MLR, RVR, SVR	0.92	[78]
he cities of Ray and Karaj, near the capital of Iran	2020	TCARI, MCARI	0.83	[79]
Anhui province ofChina	2024	$ND_{(1287,1673\ nm)}$ and GBDT	0.86	[80]
Fukang Xinjiang, China	–	MBSI + MI + FOD + KNN	0.86	This work

WC wavelet coefficients, *BPNN* BP neural network, *x-LW* x-loading weight, *WVI* water vegetation index, *GRA* grey relational analysis, *PLSR* partial least squares regression

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 22 of 24

Conclusions

In this study, the fractional order derivative and machine learning methods were applied for estimating LWC of spring wheat, and results showed the integrated Model-$_{3BSMI+FOD+KNN}$ achieved high accuracy. The two-band index and three-band index were extracted from spectrum based on FOD, and the linear regression analysis was conducted between these indices and LWC values. The fractional derivative pretreatment of spectral data enhances the implied information of the spectrum, and both spectral and these indices were closely correlated with the LWC values. The machine learning method for LWC estimation of spring wheat based on sensitive spectral and indices. The results showed the proposed method by combing two-band indices and three-band indices improved the estimation accuracy. The moisture content of plant leaves is an important indicator for measuring the water status of plants. Accurately estimating leaf moisture content is of great significance for studying plant physiology, ecology, agricultural production, and environmental protection. KNN, ANN and SWR were independently conducted to predict LWC values based on optimal combinations. The results showed that KNN performed better than ANN, SWR with higher R^2 and lower RMSE. Therefore, the results confirm that Model-$_{3BSMI+FOD+KNN}$ is significantly effective in inverting the leaf water content of spring wheat.

Author contributions
Conceptualization, Mardan Turghan; Formal analysis, Bilal Imin; Funding acquisition, Nijat kasim; Resources, Umut Hasan; Software, Mutallip Sattar; Writing-original draft, Zinhar Zununjan; Writing—review and editing, Nijat kasim and Abdugheni Abliz.

Data availability
Data available on request. The data uderlying this article will be shared on reasonable request to the corresponding author.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
This manuscript has not been published elsewhere and is not under consideration by another journal.

Competing interests
This manuscript has no competing interest to declare.

Author details
[1]School of Resources and Environment, Yili Normal University, Yining 835000, China. [2]College of Resources and Environmental Sciences, Xinjiang University, Urumqi 830046, China. [3]State Key Laboratory of Oasis and Desert Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China. [4]College of Information Management, Xinjiang University of Finance and Economics, Urumqi 830012, China.

References
1. Zhen Z, Tiansheng L, Jing C. Non-destructive estimation of winter wheat leaf moisture content using near-ground hyperspectral imaging technology. Acta Agric Scand Sect B Soil Plant Sci. 2020;70(4):294–306.
2. Feifei Y, Tao LB, Qiyuan W, Mingzhu D, Tianle Y, Dazhong L, Shijuan L, Shengping L. Rapid determination of leaf water content for monitoring waterlogging in winter wheat based on hyperspectral parameters. J Integr Agric. 2021;20(10):2613–26.
3. Fan L, Yang J, Sun X, Zhao F, Liang S, Duan D. The effects of landsat image acquisition date on winter wheat classification in the north china plain. ISPRS J Photogr Remote Sens. 2022;187:1–13.
4. Zhuang J, Li Z, Li Y, Xi Y. The timing variations and spatial pattern of winter wheat in China. In: Fifth International Conference on Agro-geoinformatics. IEEE. 2016.
5. Maneesha R, SujeetDesai S, PriyaDevi S, Gupta MJ. Estimation of crop water requirement of pineapple (*Ananas comosus* (L.) Merr.) for drip fertigation. Int J Bio-resour Stress Manage. 2002;13(1):973–80.
6. He L, Liu MR, Zhang SH, Guan HW, Wang CY, Feng W, et al. Remote estimation of leaf water concentration in winter wheat under different nitrogen treatments and plant growth stages. Precis Agric. 2002; 1–28.
7. Chipilski R, Dimitrov E, Uhr Z. Study of photosynthesis, leaf water exchange and yield of field grown common winter wheat varieties under dry prone conditions. Bulg J Agric Sci. 2022;28(5):860–5.
8. Kong W, Huang W, Ma L, Tang L, Li C, Zhou X, Casa R. Estimating vertical distribution of leaf water content within wheat canopies after head emergence. Remote Sens. 2021;13:4125.
9. Li C, Xiao Z, Liu Y, Meng X, Li X, Wang X, Li Y, Zhao C, Ren L, Yang C. Hyperspectral Estimation of Winter Wheat Leaf Water Content Based on Fractional Order Differentiation and Continuous Wavelet Transform. Agronomy. 2023;13:56.
10. Liu Y, Feng H, Yue J, Li Z, Jin X, Fan Y, Feng Z, Yang G. Estimation of above-ground biomass of potatoes based on characteristic variables extracted from UAV hyperspectral imagery. Remote Sens. 2022;14:5121.
11. Zhang J, Zhang W, Xiong S, Song Z, Tian W, Shi L, Ma X. Comparison of new hyperspectral index and machine learning models for prediction of winter wheat leaf water content. Plant Methods. 2021;17(1):1–17.
12. Kasim N, Sawut R, Abliz A, Qingdong S, Maihmuti B, Yalkun A, et al. Estimation of the relative chlorophyll content in spring wheat based on an optimized spectral index. Photogramm Eng Remote Sens. 2018;12(84):801–11.
13. Umut H, Mamat S, Shuisen C. Estimating the Leaf Area Index of Winter Wheat Based on Unmanned Aerial Vehicle RGB-Image Parameters. Sustainability. 2019;11(23):68–29.
14. Ding XL, Zhao LX, Zhou TT, Li YB, Huang XM, Zhao YL. Research on wheat leaf water content based on machine vision. Clust Comput. 2019;22(Suppl 4):9199–208.
15. Zhang J, Zhang W, Xiong S, Song Z, Ma X. Comparison of the prediction of winter wheat leaf water content by using new hyperspectral index and machine learning models. 2020.

Zununjan *et al. Plant Methods*   (2024) 20:97

Page 23 of 24

16. Wang YC, Zhang XY, Jin YT, Gu XH, Feng H, Wang C. Quantitative retrieval of water content in winter wheat leaves based on continuous wavelet transform. J Triticeae Crops. 2020;40:503–9 (**(In Chinese)**).

17. Liu SG, Hu ZQ, Han JZ, Li YY, Zhou T. Predicting grain yield and protein content of winter wheat at different growth stages by hyperspectral data integrated with growth monitor index. Comput Electron Agric. 2022;200:107–235.

18. Liu J, Ding J, Ge X, Wang J. Evaluation of total nitrogen in water via airborne hyperspectral data: potential of fractional order discretization algorithm and discrete wavelet transform analysis. Remote Sens. 2021;13(22):4643.

19. Yue JB, Feng HK, Yang GJ, Li ZH. A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. Remote Sens. 2018;10:66.

20. Yasir QM, Zhang W. Estimation of leaf water content of different leaves from different species using hyperspectral reflectance data. Ann Agric Crop Sci. 2022;7(2):111.

21. Jin X, Xu X, Song X, Li Z, Wang J, Guo W. Estimation of leaf water content in winter wheat using grey relational analysis–partial least squares modeling with hyperspectral data. Agron J. 2013;105(5):1385–92.

22. Dobrowski SZ, Pushnik JC, Zarco-Tejada PJ, Ustin SL. Simple reflectance indices track heat and water stress-induced changes in steady-state chlorophyll fluorescence at the canopy scale. Remote Sens Environ. 2005;97(3):403–14.

23. Jiashuang Q, Daxing G, Longkang N, Wen H, Yangchun R, Yuqing H. Hyperspectral inversion of leaf water content at karst and non-karst areas in north Guangxi. Guihaia. 2021;41(05):726–37.

24. Yang S, Hu L, Wu H, Ren H, Fan W. Integration of crop growth model and random forest for winter wheat yield estimation from uav hyperspectral imagery. IEEE J Sel Topics Appl Earth Observ Remote Sens. 2021;14:6253.

25. Das B, Sahoo RN, Pargal S, Krishna G, Verma R, Viswanathan C, Sehgal VK, Gupta VK. Evaluation of different water absorption bands, indices and multivariate models for water-deficit stress monitoring in rice using visible-near infrared spectroscopy. Spectrochim Acta Part A Mol Biomol Spectrosc. 2021;247:119–104.

26. Gao Y, Qiu J, Miao Y, Qiu R, Li H, Zhang M. Prediction of leaf water content in maize seedlings based on hyperspectral information. IFAC Conf Sens Contr Automat Technol Agric. 2020;52:263.

27. Sun H, Feng M, Xiao L, Yang W, Li D. Assessment of plant water status in winter wheat (*Triticum aestivum* L.) based on canopy spectral indices. PLoS ONE. 2019;14(6):e0216890.

28. Ceccato P, Flasse S, Tarantola S, Jacquemoud S, Cregoire J-M. Detecting vegetation leaf water content using reflectance in the optical domain. Remote Sens Environ. 2001;77(1):22–33.

29. Baghzouz M, Devitt D, Morris R. Assessing canopy spectral reflectance of hybrid Bermudagrass under various combinations of nitrogen and water treatments. Appl Eng Agric. 2007;23(6):763–74.

30. Wang J, Xu R, Yang S. Estimation of plant water content by spectral absorption features centered at 1450 nm and 1940 nm. Environ Monit Assess. 2009;157(1–4):459–69.

31. Wang J, Zhao C, Guo X, Huang W, Tian Q. Diagnosis of water content in wheat leaves by remote sensing method. J Agron N China. 2000;15(4):68–72 (**In Chinese**).

32. Tian Q, Gong P, Guo ZX. Feasibility analysis of using spectral reflectance to diagnose moisture status of wheat. Chin Sci Bull. 2001;45(24):2654–2650 (**In Chinese**).

33. Yi QX, Bao AM, Wang Q, Zhao J. Estimation of leaf water content in cotton by means of hyperspectral indices. Comput Electron Agric. 2013;90:144–51 (**In Chinese**).

34. Kasim N, Qingdong S, Jingzhe W, Sawut R, Nurmemet I, Isak G. Estimation of spring wheat chlorophyll content based on hyperspectral features and PLSR model. Trans Chin Soc Agric Eng (Trans CSAE). 2017;33(22):208–16 (**In Chinese**).

35. Wang J. Estimation of plant water content by spectral absorption features centered at 1,450 nm and 1,940nm regions. Environ Monit Assess. 2008.

36. Zhao S, Ding J, Ge X, Huang S, Han L. Soil salinity estimation: effects of microwave dielectric spectroscopy and important frequencies. Land Degrad Dev. 2022.

37. Zhang F, Zhou G. Deriving a light use efficiency estimation algorithm using hyperspectral and eddy covariance measurements for a maize canopy in Northeast China. Ecol Evol. 2017.

38. Mahajan GR, Das B, Gaikwad B, Murgaonkar D, et al. Monitoring properties of the salt-affected soils by multivariate analysis of the visible and near-infrared hyperspectral data. CATENA. 2021.

39. Wu C, Zheng Y, Yang H, Yang Y, Wu Z. Effects of different particle sizes on the spectral prediction of soil organic matter. CATENA. 2021.

40. Ortigueira M, Rodr'iguez-Germá L, Trujillo J. Complex Grünwald-Letnikov, Liouville, Riemann-Liouville, and Caputo derivatives for analytic functions. Commun Nonlinear Sci Numer Simul (JCR). 2011;16:4174–82.

41. Pan I, Das S. Frequency domain design of fractional order PID controller for AVR system using chaotic multi-objective optimization. Int J Electr Power Energy Syst. 2013;51:10.

42. Saadia A, Rashdi A. Fractional order integration and fuzzy logic based filter for denoising of echocardiographic image. Comput Methods Progr Biomed. 2016;137:10.

43. Hu Z, Pan C, Pan X, Zhu B. Estimation models for water content of walnut leaves based on spectral moisture index. Sci Silvae Sin. 2016;52(12):39–49.

44. Corti M, Gallina PM, Cavalli D, Cabassi G. Hyperspectral imaging of spinach canopy under combined water and nitrogen stress to estimate biomass, water, and nitrogen content. Biosys Eng. 2017;158:38–50.

45. Li X, Liu X, Liu M, Wang C, Xia X. A hyperspectral index sensitive to subtle changes in the canopy chlorophyll content under arsenic stress. Int J Appl Earth Obs Geoinf. 2015;36:41–53.

46. Zheng W, Xia Z, Bing Z, Xing L. General and specific methods studying on bands selection of hyperspectral remote sensing data. In: Geoscience and Remote Sensing Symposium, 2005, IGARSS '05. Proceedings. 2005 IEEE International. IEEE, 2005.

47. Wang L, Hunt ER Jr, Qu JJ, Hao X, Daughtry C. Towards estimation of canopy foliar biomass with spectral reflectance measurements. Remote Sens Environ. 2011;115(3):836–40.

48. Hong Y, Chen Y, Yu L, Liu Y, Liu Y, Zhang Y, Liu Y, Cheng H. Combining fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by VIS–NIR spectroscopy. Remote Sens. 2018;10:479.

49. Wang J, Ding J, Abulimiti A, Cai L. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS–NIR) spectroscopy, Ebinur Lake Wetland, Northwest China. PeerJ. 2018;6: e4703.

50. Cao Z, Cheng T, Ma X, Tian Y, Zhu Y, Yao X. A new three-band spectral index for mitigating the saturation in the estimation of leaf area index in wheat. Int J Remote Sens. 2017;38(13–14):3865–85.

51. Zhang Z, Ding J, Wang J, Ge X. Prediction of soil organic matter in northwestern china using fractional-order derivative spectroscopy and modified normalized difference indices. CATENA. 2019;185: 104257.

52. Ding J, Yang A, Wang J, Sagan V, Yu D. Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. PeerJ. 2018;6: e5714.

53. Wang J, Ding J, Li G, Liang J, Danlin Yu, Aishan T, Zhang F, Yang J, Abulimiti A, Liu J. Dynamic detection of water surface area of Ebinur Lake using multi-source satellite data (Landsat and Sentinel-1A) and its responses to changing environment. CATENA. 2019;177:189–201.

54. Inoue Y, Peñuelas J, Miyata A, Mano M. Normalized difference spectral indices for estimating photosynthetic efficiency and capacity at a canopy scale derived from hyperspectral and co 2 flux measurements in rice. Remote Sens Environ. 2008;112:156–72.

55. Gitelson AA, Gritz Y, Merzlyak MN. Relationships between leaf relative chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J Plant Physiol. 2003;160(3):271–82.

56. Jackson RD. Spectral indices in N-space. Remote Sens Environ. 1983;13(5):409–21.

57. Tian YC, Yao X, Yang J, Cao WX, Hannaway DB, Zhu Y. Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground- and space-based hyperspectral reflectance. Field Crop Res. 2011;120(2):299–310.

58. Zhang W, Zhuo Z, Peng L, Jingqi L, Sun T, Tang H, Tang J, Zhou T, Li L. Laser-induced breakdown spectroscopy for quantitative and qualitative analysis on ash fusion temperatures of coal in power plant. J Anal Atom Spectr. 2021;36:576.

59. Ye T, Dong M, Liang Y, Long J, Li W, Lu J. " Modeling and optimization of the NO generation characteristics of the coal-fired boiler based on interpretable machine learning algorithm. Int J Green Energy. 2021;51:529.

60. Zeng X-P, Luo Q, Zheng J-L, Chen G-H. An efficient neural network optimized by fruit fly optimization algorithm for user equipment association in software-defined wireless sensor network. Int J Netw Manage. 2020;30:e2135.

61. Zhang S, Wu W, Yang Z, Lin X, Ren Z, Yan Z. MLLBC: a machine learning toolbox for modelling the loss rate of the lining bearing capacity. IEEE Access. 2020;8:50256.

62. Fan S, Li Z, Xia K, Hao D. Quantitative and qualitative analysis of multicomponent gas using sensor array. Sensors. 2019;19:3917.

63. Hong Y, Shen R, Cheng H, Chen S, Liu Y. Cadmium concentration estimation in peri-urban agricultural soils: using reflectance spectroscopy, soil auxiliary information, or a combination of both ? Geoderma. 2019;354:113875.

64. Ge X, Jianli D, Xiuliang J, Jingzhe W, Xiangyue C, Xiaohang L, Jie L, Boqiang X. Estimating agricultural soil moisture content through UAV-based hyperspectral images in the arid region. Remote Sens. 2021;13(8):1562.

65. Lao C, Chen J, Zhang Z, Chen Y, Ma Y, Chen H, Gu X, Ning J, Jin J, Li X. Predicting the contents of soil salt and major water-soluble ions with fractional-order derivative spectral indices and variable selection. Comput Electron Agric. 2021;182:106031.

66. Jianing Z, Xiapeng J, Yi X, Jing M, Demei Z, Junjie W, Jingzhe W, Guofeng W. Mapping leaf chlorophyll content of mangrove forests with Sentinel-2 images of four periods. Int J Appl Earth Observ Geoinf. 2021;102:102387.

67. Angel Y, McCabe MF. Machine learning strategies for the retrieval of leaf-chlorophyll dynamics: model choice, sequential versus retraining learning, and hyperspectral predictors. Front Plant Sci. 2022;13:722442.

68. Jin XL, Xin-Gang XU, Wang JH, Xin-Chuan LI, Wang Y, Tan CW, Zhu XK. Hyperspectral estimation of leaf water content for winter wheat based on grey relational analysis (GRA). Spectrosc Spectral Anal. 2012;32(11):3103–6.

69. Sharifi A, Felegari S. Nitrogen dioxide (NO_2) pollution monitoring with sentinel-5P satellite imagery over during the coronavirus pandemic (case study: Tehran). Remote Sens Lett. 2022;13:1029–39.

70. Felegari S, Sharifi A, Khosravi M, Sabanov S. Using experimental models and multitemporal Landsat-9 images for cadmium concentration mapping. IEEE Geosci Remote Sens Lett. 2023;20:1–1. https://doi.org/10.1109/LGRS.2023.3291019.

71. Farmonov N, Amankulova K, Szatmari J, Sharifi A, Abbasi-Moghadam D, Mirhoseini N, Seyed M, Mucsi L. Crop type classification by DESIS hyperspectral imagery and machine learning algorithms. IEEE J Sel Topics Appl Earth Observ Remote Sens. 2023;16:1576.

72. Mirhoseini N, Seyed M, Abbasi-Moghadam D, Sharifi A, Farmonov N, Amankulova K, Mucsi L. Multi-spectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches. IEEE J Sel Topics Appl Earth Observ Remote Sens. 2022;16:254–66.

73. Zhao Y. Principals and method of remote sensing application and analysis. Beijing: Science Press; 2003. p. 325–418.

74. Schmitt JM. Fractional derivative analysis of diffuse reflectance spectra. Appl Spectrosc. 1998;52:840–6.

75. Tong P, Du Y, Zheng K, Wu T, Wang J. Improvement of NIR model by fractional order Savitzky-Golay derivation (FOSGD) coupled with wavelength selection. Chemometr Intell Lab Syst. 2015;143:40–8.

76. Peña-Barragán JM, Gutiérrez PA, Martínez C, Six J, Plant RE, López-Granados F. Object-based image classification of summer crops with machine learning methods. Remote Sens. 2014;6:5019–41.

77. Guo Y, Chen S, Li X, Cunha M, Senthilnath J, Cammarano D, Fu Y. Machine learning-based approaches for predicting SPAD values of maize using multi-spectral images. Remote Sens. 2022;14:1337.

78. Sharifi A, Hosseingholizadeh M. Application of Sentinel-1 data to estimate height and biomass of rice crop in Astaneh-ye Ashrafiyeh, Iran. J Indian Soc Remote Sens. 2019;48:11–9.

79. Sharifi A. Remotely sensed vegetation indices for crop nutrition mapping. J Sci Food Agric. 2020;100:5191.

80. Zhang X, Xu H, She Y, Hu C, Zhu T, Wang L, Wu L, You C, Ke J, Zhang Q, He H. Improving the prediction performance of leaf water content by coupling multi-source data with machine learning in rice (*Oryza sativa* L.). Plant Methods. 2024;20:48.

81. Esmaeili M, Dariush AM, Sharifi A, Tariq A, Li Q. Hyperspectral image band selection based on CNN embedded GA CNNeGA. IEEE J Sel Topics Appl Earth Observ Remote Sens. 2023;16:1927.

Publisher's Note