Computers and Electronics in Agriculture 193 (2022) 106728

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

# Winter wheat chlorophyll content retrieval based on machine learning using in situ hyperspectral data

Tianli Wang [a,b], Maofang Gao [a,*], Chunling Cao [b], Jiong You [c], Xiwang Zhang [d], Lanzhi Shen [e]

[a] Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture and Rural Affairs/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[b] Math School, Jilin University, Changchun 130012, China
[c] Key Laboratory of Cultivated Land Use, Ministry of Agriculture and Rural Affairs/Academy of Agricultural Planning and Engineering, MARA, Beijing 100125, China
[d] Key Laboratory of Geospatial Technology for Middle and Lower Yellow River Regions (Henan University), Ministry of Education, China
[e] Twenty First Century Aerospace Technology Co., Ltd., Beijing 100096, China

ARTICLE INFO

*Keywords:*
Gradient boosting regression trees
Nonlinear fitting
Dimensionality reduction of hyperspectral data

ABSTRACT

Chlorophyll is the basic component of plant leaves. It is closely related to the use of solar radiation, the absorption of atmospheric carbon dioxide and the process of photosynthesis. The level of chlorophyll content has a very important indication of photosynthetic efficiency and development status of plants. Because hyperspectral data has the problem of large amount of data and high redundancy, this paper deals with hyperspectral data from two aspects to establish a prediction model of leaf chlorophyll content in different growth stages of winter wheat. On the one hand, this article does not perform dimensionality reduction processing on hyperspectral data. That is, for the full-band spectral data, the Ridge regression algorithm (abbreviated as Ridge) is used to establish a multiple linear prediction model. Use the Gradient Boosting Regression Tree algorithm (abbreviated as GBRT) and the feedforward neural network based on Back Propagation algorithm (abbreviated as BP) to establish two multiple nonlinear prediction models. On the other hand, dimensionality reduction is performed on hyperspectral data. Use Elastic Net algorithm (abbreviated as EN) to reduce the dimensionality of hyperspectral data to extract sensitive bands, and then use GBRT algorithm and feedforward neural network based on Back Propagation algorithm to establish multivariate nonlinear prediction models (abbreviated as EN-GBRT, EN-BP). After the model accuracy test, the conclusions are as follows: (1) The nonlinear prediction model obtained by the gradient boosting regression tree algorithm has higher accuracy and universality than the linear model in each growth period of winter wheat ($R^2_{GBRT} > R^2_{Ridge}$). (2) The EN-GBRT prediction model has greater advantages in predicting the chlorophyll content of winter wheat at each growth stage ($R^2_{EN\text{-}GBRT} > R^2_{EN\text{-}BP}$) within two nonlinear prediction models. (3) Comprehensive comparison of the prediction accuracy of GBRT and EN-GBRT models in each growth period. In the jointing period, heading period, flowering period, filling period, milking period and maturity period of winter wheat, the accuracy of prediction using the GBRT model was higher ($R^2_{jointing} = 0.81$, $R^2_{heading} = 0.50$, $R^2_{flowering} = 0.74$, $R^2_{filling} = 0.41$, $R^2_{milking} = 0.84$, $R^2_{maturity} = 0.95$). In the whole growth period of winter wheat, the prediction accuracy of EN-GBRT model was better ($R^2_{whole} = 0.85$).

## 1. Introduction

Remote sensing technology is at the forefront of modern information technology and has the advantage of quickly and timely obtaining information on agriculture, environment and meteorology. It can obtain various quantitative information in a timely and accurate manner (Yao et al., 2014), which has been widely used in how the agricultural modernization develop. China has more people and less land, and the contradiction between food supply and demand is more prominent. Therefore, the use of hyper-spectral remote sensing technology to diagnose and monitor the growth and nutritional status of winter wheat in real-time, non-destructively, accurately and quickly (Kong et al., 2017) is of great practical significance for solving the problem of food security in the world and improving the level of agricultural production. Chlorophyll is the most important pigment in plant photosynthesis. Its content can be used to detect the nutritional status of plants and is an

* Corresponding author.
*E-mail address:* gaomaofang@caas.cn (M. Gao).

https://doi.org/10.1016/j.compag.2022.106728
Received 17 September 2021; Received in revised form 3 January 2022; Accepted 17 January 2022
Available online 24 January 2022

important indicator reflecting the overall growth status of plants (Feng et al., 2008). Therefore, nondestructive prediction of winter wheat chlorophyll content can guide agricultural production and field tillage, achieve high yield and high-quality production of wheat, and improve the level of agricultural production.

Some scholars have used the vegetation index to invert the chlorophyll content of winter wheat leaves and have made certain progress. Chunjiang et al. (2006) constructed the Canopy Chlorophyll Inversion Index and used multi-angle spectral information to retrieve the vertical distribution of crop chlorophyll through different angle conditions, reaching a significant level. Jiang et al. (2010) measured the canopy spectrum and canopy chlorophyll content of winter wheat infected with stripe rust in different growth stages and constructed nine types of vegetation indices. Among these vegetation indexes, vegetation indexes with correlation coefficients greater than 0.7 are selected as independent variables to establish an inversion model. The results show that the differential index $(D_{750}\text{-}D_{550})/(D_{750} + D_{550})$ has the best inversion accuracy and stability ($R^2 = 0.77$). Although the above research has fully excavated various vegetation indexes, the inversion of vegetation indexes does not fully retain the effective information of the spectral curve, and the inversion accuracy of the model will be affected. Therefore, it is necessary to improve the effective utilization of the spectral curve.

Researchers began to introduce more complex data processing to improve the shortcomings of a single vegetation index. Sawuti et al. (2019) used fractional differentiation method to process the hyperspectral data of wheat leaves to improve the estimation accuracy of the chlorophyll content prediction model. The results show that the vegetation index calculated by logarithmic transformation and 1.2-order differential has the best estimation accuracy ($R^2 = 0.87$). Sun et al. (2017) established a chlorophyll content estimation model using vegetation index and Back Propagation Neural Network (shorthand for BP), which has high fitting accuracy ($R^2_{Train} = 0.96$, $R^2_{Verify} = 0.70$). The above research has carried out a mathematical transformation of the hyperspectral data and improved the accuracy of the model inversion, but there is still a problem of insufficient use of the hyperspectral data information.

In order to make full use of the effective information of hyperspectral, some scholars proposed to use the curve characteristics of hyperspectral reflectance to establish an inversion model. Qian et al. (2021) used the Newton Eight-Point Interpolation method to calculate the position of the red edge of the spectral reflectance to establish a winter wheat chlorophyll content inversion model, and the inversion accuracy was good ($R^2 > 0.62$). Li et al. (2017) conducted correlation analysis on the first-order differentiation of hyperspectral data and the red-edge parameters and chlorophyll content, and selected the best red-edge parameters to establish an empirical estimation model and a neural network model, and evaluate and test them. The results show that the accuracy of the model estimated by BP neural network is significantly improved ($R^2 > 0.90$).

Furthermore, some scholars pointed out that neural networks can be used to build inversion models. Yu et al. (2019) used the PROSAIL model, continuous wavelet transformation of ground hyperspectral measured data, and combined with partial least square regression to invert winter wheat chlorophyll. The effect is better than support vector machines and artificial neural network methods. Cai et al. (2019) used Least Absolute Shrinkage and Selection Operator (LASSO) and three mainstream machine learning methods (Support Vector Machine, Random Forest (RF), Neural Network) in combination with climate and satellite data to construct various empirical models for predicting yield. The results show that the method based on machine learning is better than the regression method in crop yield modeling, and can achieve high yield prediction ($R^2 > 0.75$). Zhang et al. (2019) integrated optical, fluorescent, thermal satellite and environmental data, and used the regression method of LASSO, two machine learning methods (RF and extreme gradient boosting) and deep learning Long Short-Term Memory network (LSTM) to predict corn yield. The results show that combining multi-source data can explain more than 75% of the yield variation. The above research methods increase the utilization of hyperspectral data, and have positive significance for the application of hyperspectral technology in the non-destructive detection of chlorophyll of winter wheat leaves. At the same time, the above research also effectively uses some machine learning methods, which is of great significance to the establishment of the model. However, these research methods did not analyze the different growth periods, and there was a problem that the inversion accuracy of chlorophyll content was not high.

The Gradient Boosting Regression Tree (Pham et al., 2020) is an iterative regression tree algorithm, which consists of multiple regression trees, and the sum of the conclusions of all trees is the final answer. It is an algorithm with strong generalization ability. At present, there are few researches on the gradient boosting regression tree algorithm in inverting the chlorophyll content of winter wheat leaves. An et al. (2020) combined the rate of reflectivity change between wavelengths 'a' and 'b' ($RCRW_{a\text{-}b}$) with four machine learning methods (Gaussian Process Regression (GPR), Random Forest Regression (RFR), Support Vector Regression (SVR), and GBRT) to estimate the chlorophyll content in rice. The results of the study show that this method has achieved good results ($R^2 = 0.80$).

This paper takes winter wheat chlorophyll content as the research object, establishes an empirical regression model for different growth periods, and compares the accuracy with the measured data. On the one hand, based on the hyperspectral full-band data, the ridge regression algorithm is used to establish a linear prediction model. The feedforward neural network based on Back Propagation algorithm and the gradient boosting regression tree algorithm are used to establish nonlinear prediction models. In order to improve the prediction accuracy and the effective use of hyperspectral data, this paper uses elastic network algorithm to reduce the dimensionality of hyperspectral data. Remove data noise and redundancy, filter sensitive bands, and remove irrelevant spectral information. Selected the variable combination that contains the least redundant information and least collinearity. On the other hand, based on the hyperspectral sensitive band data, the GBRT algorithm and BP neural network are used to establish two nonlinear prediction models. Based on the results of all the above-mentioned different prediction models, the best prediction models for different growth periods of winter wheat were screened out. This paper provides a model reference for the rapid, accurate and detection of chlorophyll in winter wheat leaves, and realizes the non-destructive prediction of chlorophyll green content.

The specific objectives of this paper are as follows: (1) In the current estimation of vegetation parameter models based on hyperspectral data, there are few studies that consider the different growth periods of crops (Qu et al., 2008). Based on previous studies, this study constructs regression models of winter wheat in different growth periods. In order to obtain the best estimation model for monitoring leaf chlorophyll content under different growth periods, and provide a scientific reference for timely and dynamic grasp of the growth status of winter wheat. (2) In the current estimation of vegetation parameter models based on hyperspectral data, most of them only use part of the characteristic information of hyperspectral data, and fail to make full use of the large amount of information carried by hyperspectral data. This study analyzed the accuracy and result difference of the estimation of chlorophyll content in the whole growth period and different growth periods of winter wheat in the full spectrum band and the sensitive spectrum band. In order to improve the utilization rate of hyperspectral data, improve the estimation accuracy of hyperspectral data on the chlorophyll content of winter wheat and the ability to monitor growth, provide a reference for the use of hyperspectral data to monitor crop growth research. (3) This study fully taps the potential of the gradient boosting regression tree algorithm in estimating the chlorophyll content of winter wheat leaves, in order to improve the estimation accuracy of the model. Compare different nonlinear estimation models, and fully study the

impact of multicollinearity interference between data on model accuracy. Provide technical support for remote sensing diagnosis and prediction of chlorophyll content of winter wheat leaves, and provide ideas for using machine learning algorithms to build models.

## 2. Study area and data

### 2.1. Study area

The study area in this paper is located in Langfang City, Hebei Province, at 39°35′41′′ N and 116°35′56′′ E (Fig. 1). The area is in the mid-latitude zone, with a warm temperate continental monsoon climate (Guo et al., 2016). Summer is hot and rainy, winter is cold and dry, spring is dry and windy and sandy, and autumn is cool. The temperature is suitable and the four seasons are distinct. The annual average temperature in Langfang City is 11.9℃, the average annual frost-free period is about 183 days, and the average annual precipitation in the city is 554.9 mm. Adequate light and heat resources, rain and heat in the same season, are conducive to the growth of crops. The soil type in the study area is mainly aeolian sandy soil.

### 2.2. Data measurement and preprocessing

This experiment was carried out in the field survey in Langfang City, Hebei Province in 2019, and the six growth periods of winter wheat were measured separately. The sampling times were April 26 (the joining period), May 9 (the heading period), May 13 (the flowering period), May 22 (the grain-filling period), May 28 (the milking period), May 31 (the maturity period). The entire study is divided into 50 small areas, the sample square is 1 m × 1 m, and sample points are randomly collected in the central area. The subsequent growth period is collected in the same place to achieve repeatability.

#### 2.2.1. Ground hyperspectral data

In this study, a FieldSpec4 spectrometer (Analytical Spectral Devices, Inc., Boulder, CO, USA) was used to collect the spectral reflectance of the top second leaf of winter wheat leaves. During the acquisition process, the sensor probe always keeps vertical downwards to avoid the influence of the leaf inclination angle on the spectral reflectance. The field of view of the ASD spectrometer is 25°. According to the principle of completely covering the leaf, the collection height is selected to be 0.1–0.2 m from the vertical height of the widest part of the leaf to avoid the influence of background noise on the spectral reflectance. The spectrometer can collect spectra from 350 to 2500 nm, with a sampling interval of 1.4 nm between 350 and 1000 nm, and a sampling interval of 2 nm between 1001 and 2500 nm. The spectrometer can distinguish light with a wavelength near 700 nm with a resolution accuracy of 3 nm, and can also distinguish light with a wavelength near 1400 nm and 2100 nm, with a resolution accuracy of 10 nm.

Field monitoring of winter wheat in the study area was carried out. All spectral measurements were carried out when the weather was clear, windless and cloudless, and the measurement time was 10:00–14:00. Standard whiteboard calibration was performed before and after the measurement (the reflectance of the standard white reference board is 1, and the obtained target spectrum is the dimensionless relative reflectance). During the measurement process, set the instrument to repeat sampling five times, and the system will automatically measure five spectra at each sampling point.

This paper preprocesses the five spectral curves of each sample point measured in the study area. First delete the band affected by the absorption of atmospheric water vapor on site, and then take the average of five spectral curves as the spectral reflectance value of the sample point. The measured spectra in the field are all affected by atmospheric water vapor absorption at 1331–1480 nm, 1791–1960 nm and 2301–2500 nm, and the spectral reflectance data shows abnormal fluctuations. The spectral curves in these bands contain less useful information and are of little significance to the study of vegetation spectroscopy. At the same time, in order to facilitate subsequent data processing, this study deleted these three parts of spectral data, and processed and analyzed the remaining spectral data.

#### 2.2.2. The chlorophyll content of winter wheat leaves

In this study, SPAD502 chlorophyll meter (SPAD-502 Plus, Konica Minolta Sensing Inc., Osaka, Japan) was used to determine the chlorophyll content of winter wheat leaves. Because the correlation coefficient between SPAD value and chlorophyll content of wheat leaves of the same variety reaches a significant level, it can be used to characterize the level of chlorophyll content (Zhu et al., 2005, Luo et al., 2016a). Therefore, in this paper, the SPAD value of the leaves represents the chlorophyll content. The SPAD502 chlorophyll meter is a handheld

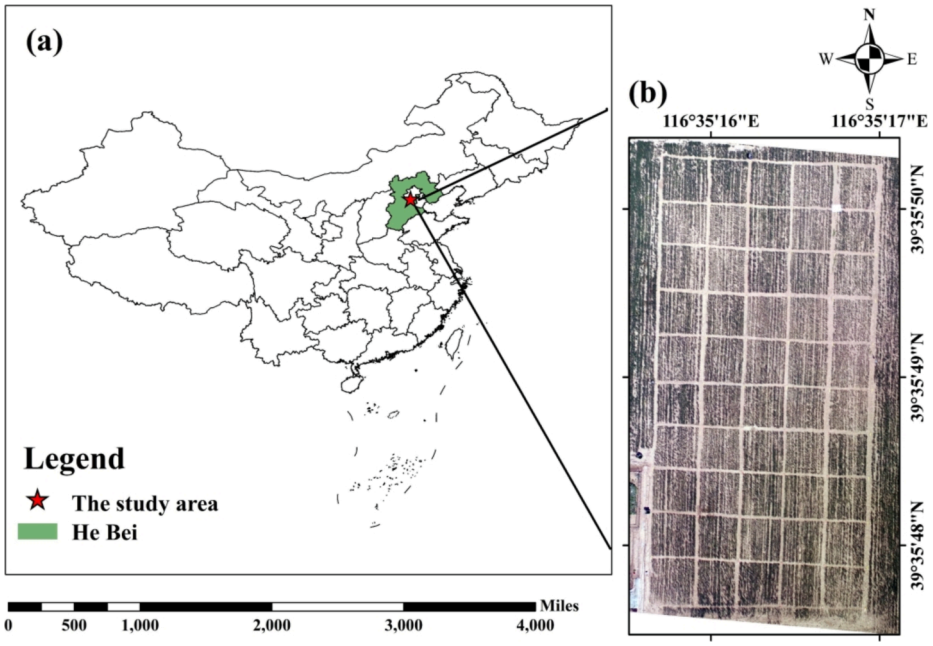


**Fig. 1.** Map of the study area: (a) The location of the Langfang study area within China's administrative region; (b) Aerial photos taken by drones in the study area.

spectrometer that can non-destructively detect the chlorophyll content of plant leaves in the field (Schlichting et al., 2015). The chlorophyll meter is calibrated every time it is turned on and used.

Top second leaf was selected in the field to measure SPAD readings at the root (60–65% from the tip of the leaf), the middle (50–55% from the tip of the leaf) and the tip (40–45% from the tip of the leaf) of the leaves (Huang et al., 2015). The readings were measured twice on both sides of the veins of each part of each leaf, for a total of six times. Avoid damaging the leaves during the measurement, and measure the SPAD value in the middle of the leaves as the chlorophyll content data of the leaves.

There are seven data sets in this study, which are the jointing stage, the heading stage, the flowering stage, the grain-filling stage, the milking stage, the maturity stage and the whole grow phase. Table 1 shows the sample size of the seven data sets and the statistical information about the chlorophyll content in each growth period. From Table 1, we can clearly see the changing trend of chlorophyll content. From the jointing stage to the grain-filling stage, as the winter wheat grows and develops, the chlorophyll content of winter wheat leaves increases continuously. From the grain-filling stage to the maturity stage, the nutrition of the wheat plant is mainly used for reproductive growth, the structure of the leaves is changed, the chlorophyll is gradually disintegrated, and the chlorophyll content of the leaves is continuously reduced.

A 10-fold cross-validation is established to verify the prediction performance of the prediction model. Divide the data set into ten parts and take turns using nine parts as training data and one part as validation data. In this paper, ten times of 10-fold cross-validation are performed, and then the average value is calculated as an estimate of the accuracy of the prediction model.

## 3. Methodology

### 3.1. Analytical framework

Fig. 2 shows the workflow used to retrieve the chlorophyll content of winter wheat leaves in this study. First, the in-situ hyperspectral data is preprocessed to obtain dimensionality reduction sensitive band and full band data. Secondly, the winter wheat SPAD database is divided into training set and validation set. Third, establish five machine learning models. Construct Ridge linear prediction model, BP nonlinear prediction model, GBRT nonlinear prediction model, EN-GBRT nonlinear prediction model and EN-BP nonlinear prediction model. Finally, by comparing the model determination coefficient and the root mean square error, the performance of the five models is evaluated and compared. Sections 3.2–3.5 show detailed information.

### 3.2. Regression algorithm

#### 3.2.1. Ridge regression algorithm

In 1970, Hoerl and Kennard (Hoerl and Kennard, 1970) proposed the Ridge Regression algorithm (Shorthand for: Ridge). Since then, the ridge regression estimation method has been widely used in practice. It is a biased estimate, an improvement to the least square estimation, and its function is to reduce the fit of the model and prevent over-fitting.

Compared with ordinary linear regression, Ridge Regression adds the L2 regularization term. For the linear regression model $y = X\beta + \varepsilon$, the objective function of the regression parameter $\beta$ is:

$$\widehat{\beta} = argmin_{\beta}\frac{1}{n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2 \tag{1}$$

where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$. y is the dependent variable of the prediction model, that is, the chlorophyll content value of each sample point. X is the independent variable of the prediction model, that is, the spectral reflectance value of each sample point. β is the regression coefficient and ε is the error. λ is a non-negative number, used to balance the variance and bias of the model, and the selection of λ is determined by the method of generalized cross-validation minimization. As λ increases, the variance of the model decreases and the deviation increases. The estimation $\widehat{\beta}$ as β is more stable than the least square estimation.

Bias is an important characteristic of ridge regression. Ridge regression analysis is a biased estimation method specially used for collinearity data analysis. This method is to give up the unbiasedness and partial accuracy of the least squares' method, and seek a regression process that is more realistic.

#### 3.2.2. Gradient boosting regression trees algorithm

Gradient Boosting Regression Trees (Friedman, 2001) is an iterative regression tree algorithm, which consists of multiple regression trees. With regression tree as the base learner, each regression tree is generated on the basis of the previous regression tree. The whole algorithm is learning and optimizing according to the forward optimization algorithm. The final result of GBRT is to add the results of all base learners. The specific steps of the algorithm are as follows:

The input is a training set sample $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, Maximum number of iterations is T, loss function is L. The output is a strong learner $f(x)$.

Step 1: Initialize the weak learner:

$$f_0(x) = \underbrace{argmin}_{c}\sum_{i=1}^{m} L(y_i, c) \tag{2}$$

Step 2: For the number of iterations $t = 1, 2, \cdots, T$, there are:

a) For samples $i = 1, 2, \cdots, m$, calculate the negative gradient

$$r_{t_i} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{t-1}(x)} \tag{3}$$

b) Use $(x_i, r_{t_i})(i = 1, 2, \cdots, m)$ to fit a CART regression tree. Get the t-th regression tree, and its corresponding leaf node area is $R_{t_j}, j = 1, 2, \cdots, J$. Where J is the number of leaf nodes of the regression tree t.

c) For the leaf area $j = 1, 2, \cdots, J$, calculate the best fit value:

$$c_{t_j} = \underbrace{argmin}_{c}\sum_{x_i \in R_{t_j}} L(y_i, f_{t-1}(x_i) + c) \tag{4}$$

d) Update strong learner:

$$f_t(x) = f_{t-1}(x) + \sum_{j=1}^{J} c_{t_j} I(x \in R_{t_j}) \tag{5}$$

**Table 1**
The number of winter wheat samples in the seven growth periods and the statistical information (including the maximum, minimum and average values) of the chlorophyll content of winter wheat leaves in each growth period.

| Seven growth periods | The number of winter wheat samples | Maximum | Minimum | Average |
|---|---|---|---|---|
| Jointing | 115 | 56.80 | 25.20 | 44.07 |
| Heading | 106 | 68.90 | 45.90 | 55.01 |
| Flowering | 63 | 65.10 | 46.40 | 53.97 |
| Filling | 100 | 76.80 | 29.20 | 54.99 |
| Milking | 132 | 65.30 | 17.30 | 49.62 |
| Maturity | 150 | 63.60 | 2.50 | 41.28 |
| Whole | 666 | 76.80 | 2.50 | 48.64 |

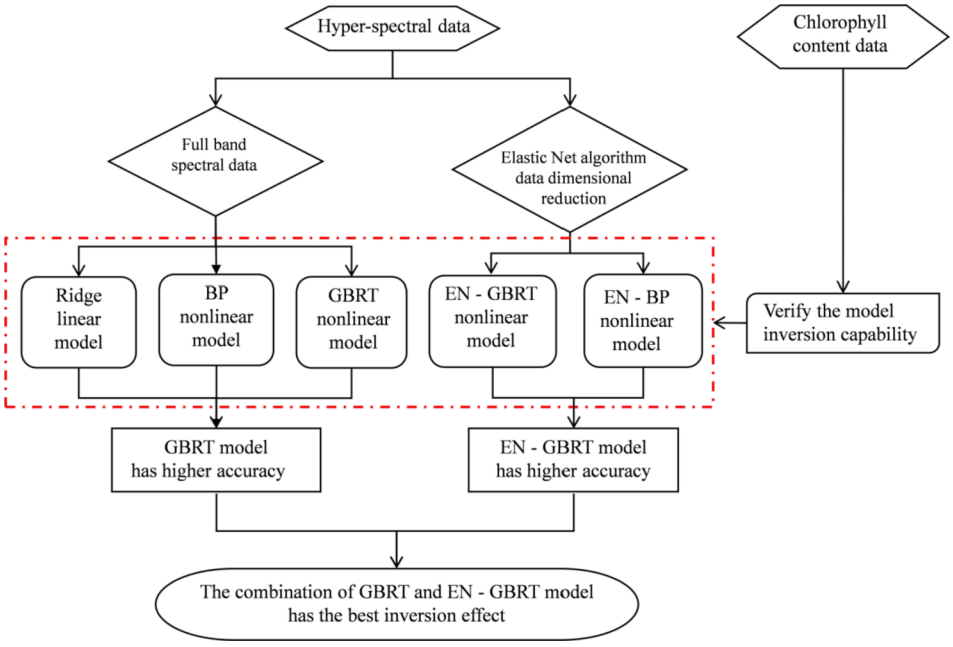


**Fig. 2.** Construction of a workflow for prediction of chlorophyll content in winter wheat leaves.

Step 3: Obtain the expression of the strong learner $f(x)$:

$$f(x) = f_T(x) = f_0(x) + \sum_{t=1}^{T}\sum_{j=1}^{J} c_{t_j} I\left(x \in R_{t_j}\right) \tag{6}$$

### 3.2.3. Back propagation neural network algorithm

The feedforward neural network based on Back Propagation algorithm (abbreviated as BP) (Robert and Hecht-Nielsen, 1988) algorithm: Use gradient descent to search the hypothesis space of possible weight vectors to find the weight vector of the best fitting example. Specifically, the loss function is used to move to the negative gradient direction of the loss function every time until the loss function reaches the minimum value. The backpropagation algorithm is based on the loss function to find the partial derivative of the weight and bias of each layer of the loss function, also known as the gradient. Use this value to update the initial weights and bias terms, until the loss function achieves the minimum value or the set number of iterations is completed (Srivastava et al., 2021). In this way, the optimal parameters in the neural network are calculated. The steps of the BP neural network algorithm are as follows (Chen et al., 2007):

Step 1: Initialize the weights and bias items in the network, denoted as $w^{(0)}, b_1^{(0)}, v^{(0)}, b_2^{(0)}$.

Step 2: Activate the forward propagation to obtain the expected value of the output and loss function of each layer:

$$E(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{7}$$

where $\theta$ represents the parameter set, $y$ represents the true value, $\widehat{y}$ represents the predicted value, and $\frac{1}{n}$ represents the average of the total error value.

Step 3: According to the loss function, calculate the error term of the output unit and the error term of the hidden unit. The error term of the output unit is the gradient value or partial derivative of the loss function with respect to the output unit. According to the chain rule:

$$\nabla_{(k)} v = \frac{\partial E}{\partial v} = \frac{\partial net_2}{\partial v}\frac{\partial \widehat{y}}{\partial net_2}\frac{\partial E}{\partial \widehat{y}} \tag{8}$$

$$\nabla_{(k)} b_2 = \frac{\partial E}{\partial b_2} = \frac{\partial net_2}{\partial b_2}\frac{\partial \widehat{y}}{\partial net_2}\frac{\partial E}{\partial \widehat{y}} \tag{9}$$

Step 4: Calculate the gradient value or partial derivative of the loss function with respect to the hidden unit. According to the chain rule:

$$\nabla_{(k)} w = \frac{\partial E}{\partial w} = \frac{\partial net_1}{\partial w}\frac{\partial h}{\partial net_1}\frac{\partial net_2}{\partial h}\frac{\partial \widehat{y}}{\partial net_2}\frac{\partial E}{\partial \widehat{y}} \tag{10}$$

$$\nabla_{(k)} b_1 = \frac{\partial E}{\partial b_1} = \frac{\partial net_1}{\partial b_1}\frac{\partial h}{\partial net_1}\frac{\partial net_2}{\partial h}\frac{\partial \widehat{y}}{\partial net_2}\frac{\partial E}{\partial \widehat{y}} \tag{11}$$

Step 5: Update the weight and bias item output unit parameter update in the neural network:

$$v^{(k)} = v^{(k-1)} - \eta\nabla_{(k)} v = v^{(k-1)} - \eta\frac{\partial E}{\partial v}, b_2^{(k)} = b_2^{(k-1)} - \eta\frac{\partial E}{\partial b_2} \tag{12}$$

Hidden unit parameter update:

$$w^{(k)} = w^{(k-1)} - \eta\nabla_{(k)} w = w^{(k-1)} - \eta\frac{\partial E}{\partial w}, b_1^{(k)} = b_1^{(k-1)} - \eta\frac{\partial E}{\partial b_1} \tag{13}$$

where $\eta$ represents the learning rate, and $k = 1, 2, \cdots, n$ represents the number of updates or the number of iterations.

Step 6: Repeat steps 2–4 until the loss function is less than the predetermined threshold or the number of iterations is used up. The output parameters at this time are the current best parameters.

## 3.3. Full-band prediction of hyperspectral data

### 3.3.1. Linear prediction model—Ridge regression

Ridge regression is essentially a linear regression method. This method only adds the restriction of regularization when establishing the regression equation, so as to achieve the effect of solving over-fitting (Macedo, 2017). Add the L2 norm to the loss function as a penalty term. Equation (1) is equivalent to $\min_{\alpha}\sum_{i=1}^{m}(x_i\alpha - y_i)^2 + \sum_{i=1}^{m}\lambda\alpha^2$. Among them, the greater the λ value, the greater the penalty for the entire loss function. Therefore, determining the appropriate value of λ has become the key to the ridge regression method.

(1) The ridge trace method determines the value of λ

From $\widehat{\beta} = (X^TX + \lambda I)^{-1}X^Ty$, $\widehat{\beta}$ is a function of λ. When λ ∈ [0, ∞),

the$\widehat{\beta}-\lambda$ curve in the plane Cartesian coordinate system is called a ridge trace curve (Huang and Niu, 2011). The point where $\widehat{\beta}$ tends to stabilize is the best λ value to be found, as shown in Fig. 3 (taking the maturity stage as an example).

(2) 10-fold cross-validation method to determine the value of λ

10-fold cross-validation (Fushiki, 2011) is to split the data set into 10 data groups (the sample size of each group is roughly the same), 9 groups are randomly selected from the 10 groups for model training, and the remaining 1 group is used for model testing. Therefore, each set of training set and validation set will correspond to a model and model score (such as mean square error), and then an average score can be obtained. In this paper, the λ value with the highest average score is selected as the optimal λ value for ridge regression modeling.

#### 3.3.2. Non-linear prediction model—GBRT regression

GBRT model is an iterative regression tree algorithm (Liu et al., 2016). The algorithm consists of multiple regression trees, and the sum of the conclusions of all trees is the final answer. It is an algorithm with strong generalization ability. Tuning parameters in machine learning algorithms is an important task. Among them, the parameters that need to be adjusted for GBRT mainly include: n estimators, max depth, subsample, learning rate, alpha. Taking the maturity stage as an example, as shown in Fig. 4, the value of each parameter with the highest score is selected as the parameter of the model to make the model achieve the best effect.

The data sets of seven growth periods were adjusted according to the above process to obtain the best prediction accuracy of the GBRT model in each growth period.

### 3.4. Sensitive band prediction of hyperspectral data

Hyperspectral data is rich in information, with many bands, and there is a high correlation between the bands. This leads to the existence of multiple collinearities between spectral data, and the redundancy of spectral information increases (Atzberger et al., 2010). The existence of multicollinearity between variables will greatly reduce the accuracy of the model prediction and the stability, which will reduce the accuracy of the prediction results. Therefore, this paper considers applying hyperspectral data to a nonlinear empirical regression model after dimensionality reduction.

#### 3.4.1. EN dimensionality reduction

The EN dimensionality reduction algorithm (Zou and Hastie, 2005) is not limited by the number of samples. On the basis of the multiple linear regression equation, penalty terms of convex combination of L1 and L2 are added, and features that have little influence on dependent variables are deleted, so as to achieve the purpose of variable screening. The objective function is as follows:

$$J(\beta)=\sum(y-X\beta)^2+\lambda\sum\left(\alpha|\beta|+(1-\alpha)\beta^2\right) \tag{14}$$

where the selection of α is based on the criterion of the minimum mean square errors of the training sample and of the prediction bias, where λ is a non-negative number used to balance the variance and bias of the model, and the selection of λ is determined by the generalized cross-validation minimization method.

In this paper, α value is selected according to the principle of minimum mean square error (MSE) of training samples. Due to the $\alpha\in(0,1)$, the interval (0,1) was divided into 10 parts. The MSE values corresponding to different α values and the number of bands were shown in Fig. 5 (taking the whole growth period as an example). The larger the alpha value is, the higher the weight of L1 regularization in the EN algorithm will be, and the fewer bands will be selected. The value corresponding to the smallest MSE in the experiment is 0.4, so $\alpha=0.4$ in the final dimensionality reduction model of EN algorithm.

When $\alpha=0.4$, the EN algorithm has the best dimensionality reduction effect. In the EN algorithm, the parameter λ value ($\lambda=1.89\times10^{-3}$) and the number of bands ($n=784$) are determined when MSE takes the minimum value (MSE = 19.05). As shown in Fig. 6.

It is finally determined that 784 sensitive bands can be obtained after dimensionality reduction using the EN algorithm (as shown in Fig. 7). It can be seen from the figure that the selected bands are basically distributed at each inflection point of the spectral curve, and contain most of the effective information of the hyperspectral data.

#### 3.4.2. Nonlinear prediction models — EN-GBRT, EN-BP

The EN dimension reduction algorithm was used to optimize the hyperspectral data for the first time, and then the GBRT algorithm and BP neural network algorithm were used for model combination and secondary optimization respectively. Results two nonlinear prediction models EN-GBRT and EN-BP were obtained. The steps to adjust the parameters are as described above.

### 3.5. Model evaluation method

In this paper, coefficient of determination ($R^2$; Equation (15)) and Root Mean Square Error (RMSE; Equation (16)) were used as indexes to evaluate the prediction ability of each model. The calculation formula is:

$$R^2=\left[\frac{\sum(y_i-\overline{y_i})(y_j-\overline{y_j})}{\sqrt{\sum(y_i-\overline{y_i})^2(y_j-\overline{y_j})^2}}\right]^2 \tag{15}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y_i})^2} \tag{16}$$

where n is the number of samples,$y_i$ and $y_j$ represent the measured value and predicted value, respectively. And $\overline{y_i}$ and $\overline{y_j}$ represent the average value of the measured value and predicted value, respectively.

$R^2$ is the model determination coefficient. The higher the value, the stronger the predictive ability of the model. RMSE is the root mean square error of the model. The smaller the value, the smaller the deviation between the model's measured value and the predicted value.

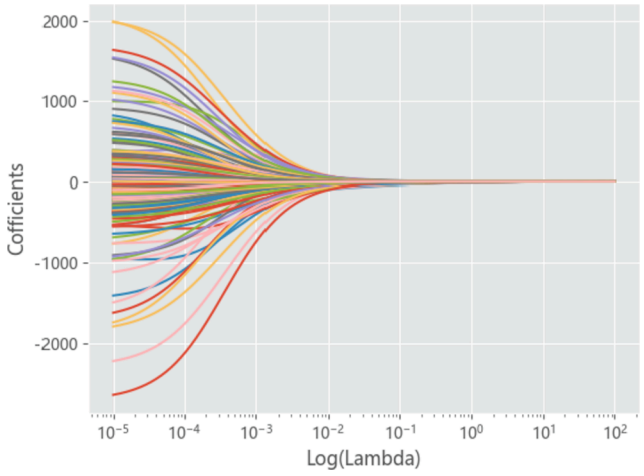


**Fig. 3.** Ridge trace method is used to determine the best λ value. The point at which $\widehat{\beta}$ tends to be stable is the best λ value to be sought (take the maturity stage as an example).

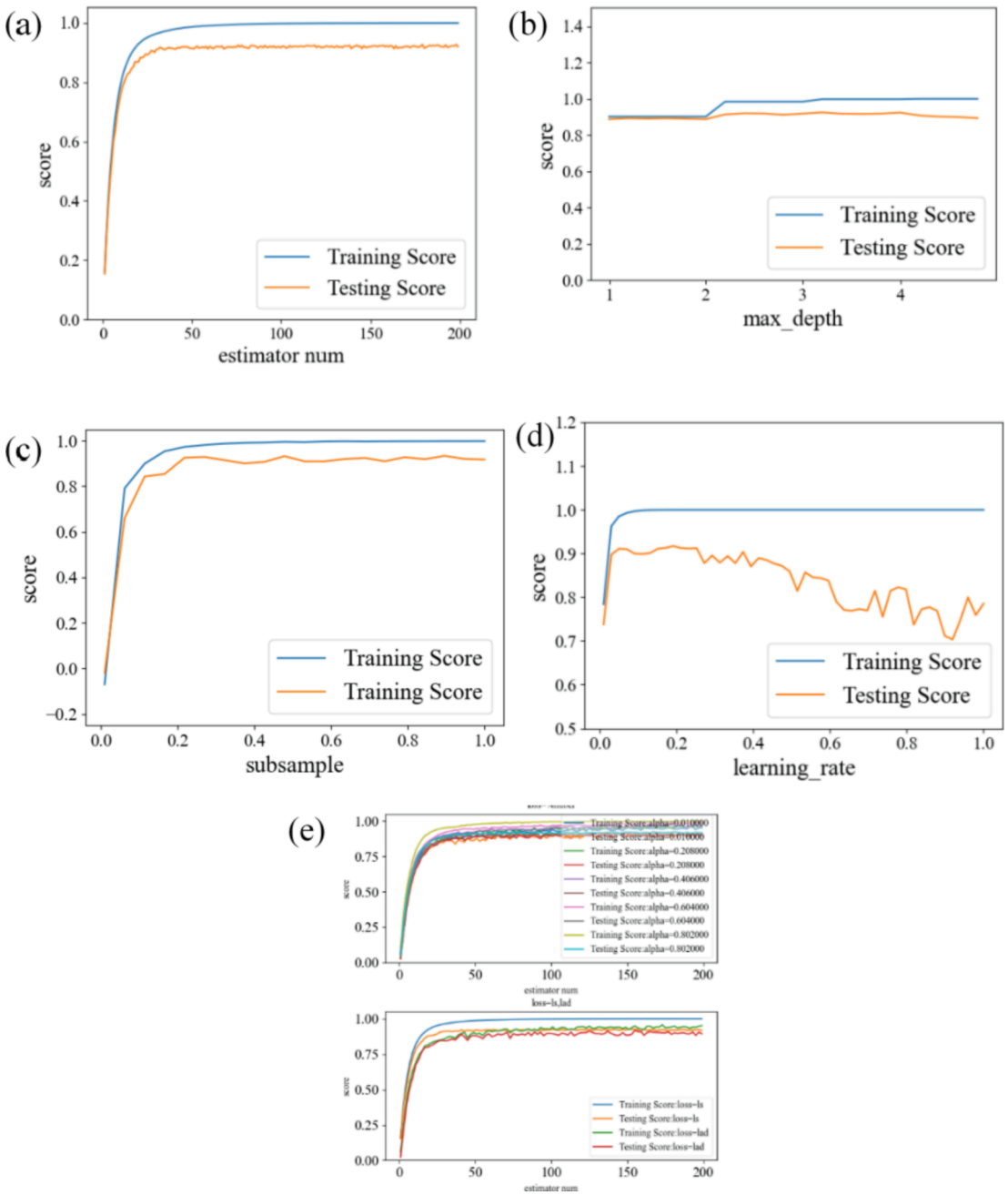


**Fig. 4.** The maturity period was taken as an example to adjust the parameters, and the value with the highest score was selected as the parameter of the model, so that the model could achieve the optimal effect. (a) Select n estimators' parameter, (b) max depth parameter, (c) Subsample parameter, (d) learning rate parameter, (e) Alpha parameter and kernel function.

## 4. Results

### 4.1. Hyperspectral full-band data prediction results

#### 4.1.1. Ridge regression model

In this paper, a 10-fold cross-validation method was used to determine the $\lambda$ value at different growth stages. The $\lambda$ was 1.61 at jointing stage, 2.83 at heading stage, 2.41 at flowering stage, 10.35 at grain-filling stage, 0.04 at milking stage, $3 \times 10^{-3}$ at maturity stage, and $9 \times 10^{-4}$ at whole growth stage. Based on the $\lambda$ value above, the ridge regression prediction model for different growth periods was improved, and the accuracy of the validation set for different growth periods was obtained from the model prediction, as shown in Fig. 8. In the jointing stage, milking stage and maturity stage, the prediction accuracy is better ($R^2_{Jointing} = 0.73$, $R^2_{Milking} = 0.70$, $R^2_{Maturity} = 0.89$, $R^2_{Whole} = 0.83$). The prediction accuracy in the flowering stage is great ($R^2_{Flowering} = 0.67$), the prediction accuracy is poor in the heading stage and the filling stage ($R^2_{Heading} = 0.45$, $R^2_{Filling} = 0.29$).

The $R^2$ and RMSE of all growth periods are compared. The results show that the prediction accuracy during the filling period is the lowest ($R^2 = 0.29$, RMSE = 3.37), the highest prediction accuracy in the maturity period ($R^2 = 0.89$, RMSE = 3.89). The prediction accuracy of heading stage and grain-filling stage is lower ($R^2 < 0.5$), mainly because the difference of chlorophyll content in these two stages is small, the fluctuation of sample data is small, and the sample data are scattered around the regression line, so the model determination coefficient is low. On the contrary, in the remaining growth period, the sample data are concentrated around the regression line and present a certain

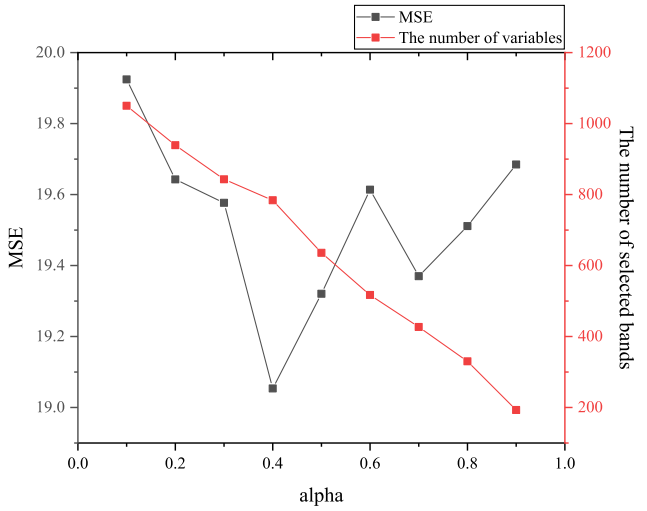


**Fig. 5.** Mean square error (MSE) values and the number of bands under different alpha values.

gradient, so the model determination coefficient is great ($R^2 > 0.6$).

#### 4.1.2. BP model

For the BP neural network model, this research makes the following design. First, select a three-layer neural network, set the neuron activation functions of the hidden layer and output layer to "logsig" and "purelin", and select "trailingdm" for the training function. Second, "dropout" is introduced to prevent the model from overfitting (Srivastava et al., 2014). Set the minimum error of the training target to 0.001, the additional momentum factor is set to 0.9, and the learning rate to 0.05. Finally, after many experiments to determine the best training parameters for different growth periods, the results are shown in Table 2.

According to the results in Table 2, in the jointing stage, heading stage, filling stage, maturity stage and whole grow stage, the prediction effect of the BP neural network is better than the Ridge model. In the flowering period and the milking period, although the $R^2$ of the two is the same, the RMSE of the BP neural network is lower. Based on the above comparison, it shows that the effect of the BP nonlinear estimation model is better than the Ridge linear model.

#### 4.1.3. GBRT model

This paper divides the training set and the validation set through 10-fold cross-validation. The chlorophyll content of leaves in different growth stages of winter wheat was retrieved by the GBRT model, and the prediction result of the verification set is shown in Fig. 9. In the jointing stage, flowering stage, milking stage and maturity stage, the prediction accuracy is better ($R^2_{Jointing} = 0.81$, $R^2_{Flowering} = 0.74$, $R^2_{Milking} = 0.84$, $R^2_{Maturity} = 0.95$, $R^2_{Whole} = 0.84$). The prediction accuracy is poor in the heading stage and the filling stage ($R^2_{Heading} = 0.50$, $R^2_{Filling} = 0.41$). It can be seen from the figure that the fitting effect of heading stage and filling stage is also poor ($R^2 < 0.5$), which verifies the above reasons. Small difference in sample data will lead to low model determination coefficient.

The results show that the prediction accuracy during the filling period is the lowest ($R^2 = 0.41$, RMSE = 3.06), the highest prediction accuracy in the maturity period ($R^2 = 0.95$, RMSE = 3.64). In general, the model has a better prediction accuracy in the seven reproductive stages.

#### 4.1.4. Comparison of prediction results of linear and nonlinear models

The prediction results of the above-mentioned linear prediction model — Ridge and the nonlinear prediction model — GBRT and BP for different growth periods are compared and analyzed, as shown in Table 3. The results show that the prediction results of GBRT and BP are all higher than that of ridge regression ($R^2_{GBRT} > R^2_{Ridge}$, $R^2_{BP} \geq R^2_{Ridge}$). Therefore, the prediction ability of the nonlinear model in this article is better than that of the linear model.

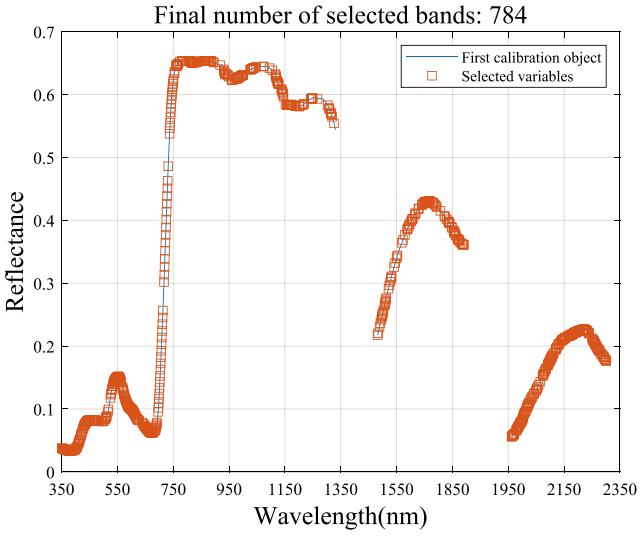


**Fig. 7.** The dimensionality reduction results of the whole growth period obtained by using Elastic Net algorithm, including the distribution of 784 bands.

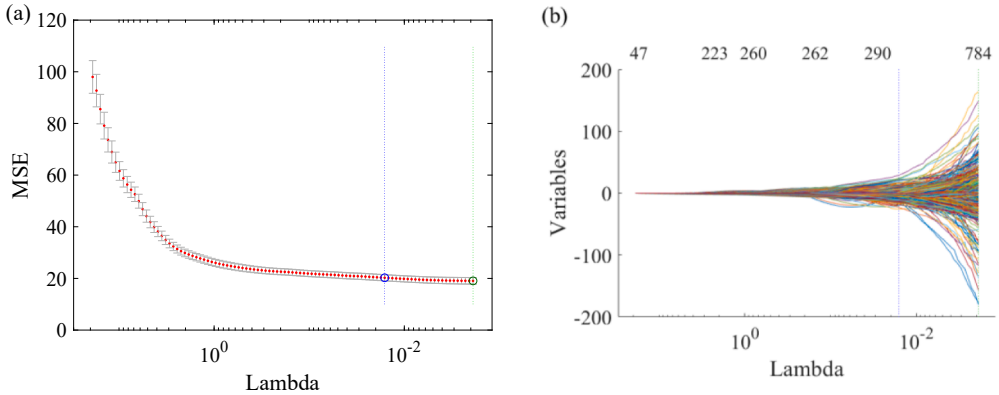


**Fig. 6.** (a) In the Elastic Net algorithm, cross-validation is used to obtain the λ value ($\lambda = 1.89 \times 10^{-3}$) corresponding to the value with the smallest mean square error (MSE = 19.05). (b) The number on the upper side is the number of bands corresponding to the λ value. Determine the number of selected bands (n = 784) and the corresponding coefficients according to the λ value obtained in (a).

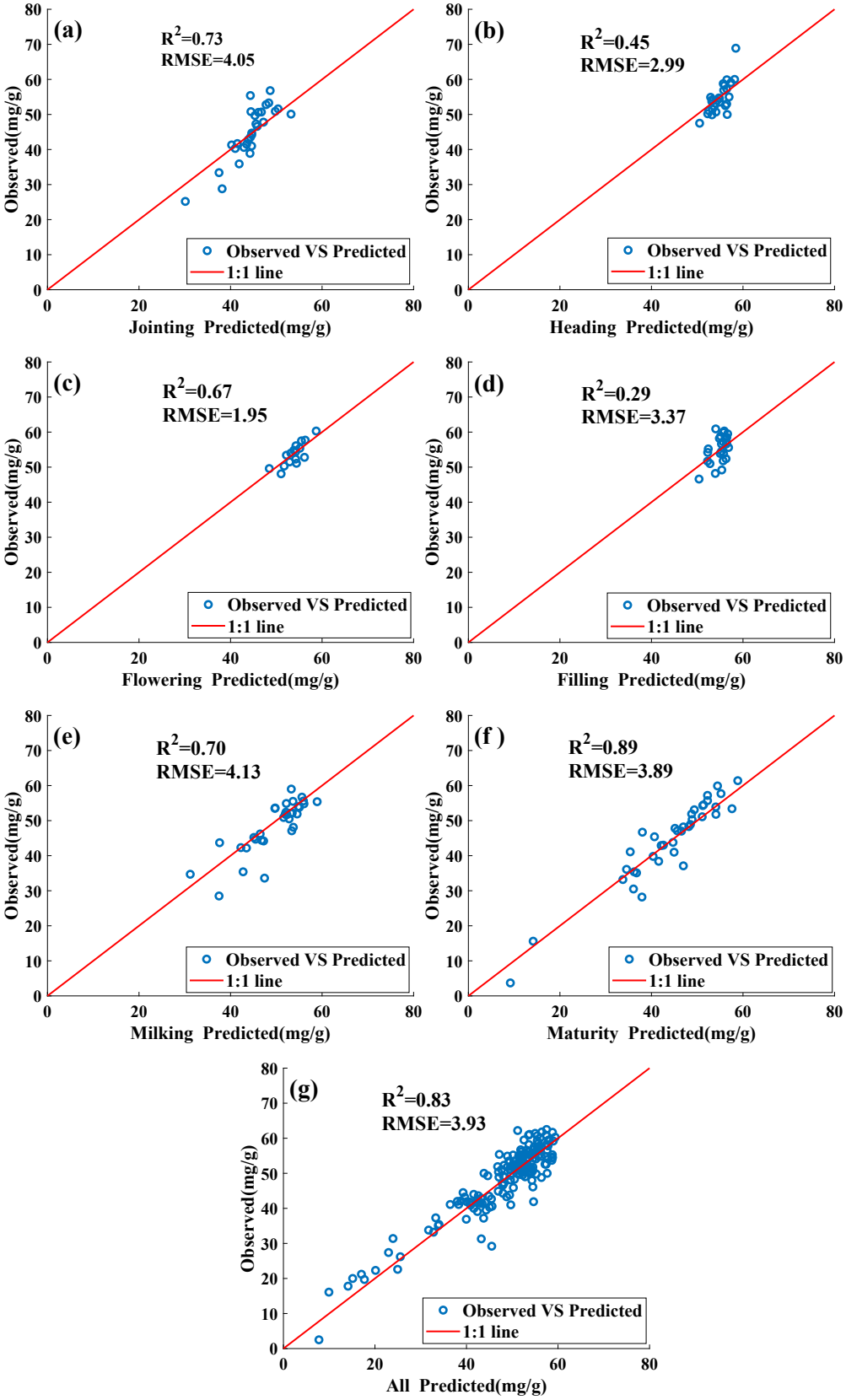


**Fig. 8.** The prediction accuracy of chlorophyll content of winter wheat based on ridge regression model for seven growth periods. The prediction results of the training set of the seven growth periods are respectively (a) (b) (c) (d) (e) (f) (g). Among them (a) is the jointing period, (b) is the heading period, (c) is the flowering period, (d) is the filling period, (e) is the milk maturity period, (f) is the maturity period, and (g) is the whole grow period.

**Table 2**
Detailed design information and results of the BP neural network estimation model at different growth stages.

| Growth stage | Maximum training times | Proportion of dropout | Number of neurons | $R^2_{BP}$ | $RMSE_{BP}$ |
|---|---|---|---|---|---|
| Jointing | 3700 | 0.38 | 18:5:1 | 0.75 | 3.85 |
| Heading | 3600 | 0.58 | 18:6:1 | 0.46 | 2.81 |
| Flowering | 4000 | 0.78 | 24:5:1 | 0.67 | 1.56 |
| Filling | 3000 | 0.73 | 18:5:1 | 0.32 | 2.34 |
| Milking | 3000 | 0.58 | 19:8:1 | 0.70 | 3.76 |
| Maturity | 7000 | 0.42 | 25:7:1 | 0.91 | 4.91 |
| Whole | 6000 | 0.35 | 21:5:1 | 0.84 | 4.29 |

The changing trends of the estimation ability of the three models at different growth stages are basically the same, indicating that the sample data of the heading and filling stage limits the estimation ability of the model, rather than the abnormality of the estimation model. At the same time, it can be seen that the accuracy of the GBRT nonlinear estimation model is higher than the Ridge linear estimation model at each growth stage. During the whole grow period, the improvement is the least, only 1%. However, the improvement was the most in the milking stage, an increase of 14%. Although the BP neural network nonlinear model has higher prediction accuracy than the Ridge linear model, it is not as effective as the GBRT nonlinear model.

In summary, the nonlinear prediction model is more effective in the process of inverting the chlorophyll content of winter wheat. Because of the interference of multicollinearity between the data in the linear prediction model, the prediction effect of the model is not as good as that of the nonlinear model.

### 4.2. Prediction results of hyperspectral sensitive band data

#### 4.2.1. EN-GBRT model

After the dimensionality of the hyperspectral data is reduced, the data set is divided by ten-fold cross-validation, and the data set with the least redundancy is obtained. As mentioned above, the optimal parameters were re-selected in different growth stages to improve the nonlinear prediction model of EN-GBRT. The leaf chlorophyll content of winter wheat at different growth stages was retrieved by the EN-GBRT model, and the results were shown in Fig. 10. In the jointing stage, flowering stage, milking stage and maturity stage, the prediction accuracy is better ($R^2_{Jointing} = 0.81$, $R^2_{Flowering} = 0.69$, $R^2_{Milking} = 0.77$, $R^2_{Maturity} = 0.94$, $R^2_{Whole} = 0.85$). The prediction accuracy is poor in the heading stage and the filling stage ($R^2_{Heading} = 0.49$, $R^2_{Filling} = 0.41$).

The prediction result of the GBRT model and the EN-GBRT model are compared. As shown in Table 4, when inverting the chlorophyll content of winter wheat at different growth periods, the GBRT model has a better prediction accuracy ($R^2_{EN\text{-}GBRT} \leq R^2_{GBRT}$). The EN-GBRT model has a better prediction accuracy when inverting the whole growth period ($R^2_{EN\text{-}GBRT} = 0.85 > R^2_{GBRT} = 0.84$, $RMSE_{EN\text{-}GBRT} = 4.13 < RMSE_{GBRT} = 4.18$) and the amount of data is large.

It can be seen from the table that the prediction capabilities of GBRT model and EN-GBRT model are almost the same, and the biggest difference is at the milking stage, which is only 7%. In the jointing stage, jointing stage, filling stage, mature stage and full growth stage, the prediction accuracy of the GBRT model and the EN-GBRT model are basically the same. This indicates that the hyperspectral sensitive band obtained by the EN dimension reduction algorithm can fully represent the whole band spectrum and effectively retain all the features of the original spectrum.

In general, the GBRT model and the EN-GBRT model have strong prediction capabilities. If you need to invert the chlorophyll content of winter wheat at different growth periods, the GBRT model is more appropriate and the prediction accuracy is better. If only the chlorophyll content of the whole growth period of winter wheat needs to be inverted, the EN-GBRT model is the best choice.

#### 4.2.2. EN-BP model

After reducing the dimensionality of the hyperspectral data, the chlorophyll content of leaves at different growth stages of winter wheat was estimated by the BP neural network algorithm, and the results are shown in Fig. 11. This study chooses a 3-layer neural network, including 2 hidden layers and an output layer to design the EN-BP model. The hidden layer uses the "logsig" neuron activation function, the output layer uses the "purelin" neuron activation function, and the training function uses "trailingdm". The number of hidden layers and neurons is obtained through continuous experiments, starting with a small value, such as one hidden layer and one neuron. If the model is under-fitting, slowly add more layers and neurons, if over-fitting, reduce the number of layers and neurons. The training strategy settings are different for different reproductive periods, and the specific settings are shown in Table 5. The minimum error of the training target is set to 0.001, the additional momentum factor is set to 0.90 and the learning rate is set to 0.05.

Comparing the EN-GBRT model with the EN-BP model, as shown in Fig. 11. Although the prediction accuracy of the validation set of the EN-BP model is not as good as the EN-GBRT model ($R^2_{EN\text{-}GBRT} \geq R^2_{EN\text{-}BP}$, $RMSE_{EN\text{-}GBRT} < RMSE_{EN\text{-}BP}$), the difference between the two models is small, both within 4%.

### 4.3. The best estimation model for different growth periods

The prediction results of the above five models of Ridge, GBRT, BP, EN-GBRT, and EN-BP are compared and analyzed.

1. The prediction accuracy of the ridge regression model in the seven growth periods is lower than the GBRT model, BP neural network model, EN-GBRT model and EN-BP model. This is due to the multicollinearity interference between the independent variable data and the redundancy between the data, which will cause the model's prediction accuracy to decrease. Collinearity is an approximately linear relationship between independent variables, and this relationship will have a great impact on regression analysis. Because regression analysis needs to understand the relationship between each input variable and output variable, high collinearity means that there is a certain functional relationship between the independent variables, so there is no way to fix other conditions to analyze the influence of a single input variable on the output variable. This causes analysis errors, so the effect of high collinearity needs to be excluded during regression analysis. In the GBRT algorithm, the fitting direction of each new decision tree is the negative gradient direction between the results of all previous decision trees and the target. This algorithm can better deal with the non-linear relationship between multivariate and target, and the calculation efficiency is higher. In addition, GBDT rarely has over-fitting problems (Natekin and Knoll, 2013). Therefore, the prediction accuracy of the nonlinear model is higher.
2. The prediction accuracy of EN-BP model in seven growth periods is higher than that of BP neural network model. This is due to the multicollinearity interference between input neurons, which is the main reason for the decrease of the prediction performance of the BP neural network. Therefore, the prediction effect of the EN-BP model after using the elastic net algorithm to reduce the redundancy between the data is better, and the prediction accuracy during the jointing and flowering periods is better than that of the EN-GBRT model ($R^2_{EN\text{-}GBRT} = R^2_{EN\text{-}BP}$, $RMSE_{EN\text{-}GBRT} > RMSE_{EN\text{-}BP}$).
3. Among the five models, the GBRT and EN-GBRT prediction models have relatively good results, at the same time, the prediction accuracy in different growth periods is also comparable. In the jointing stage, heading stage, filling stage, maturity stage and the whole growth stage, the difference between the prediction accuracy of the

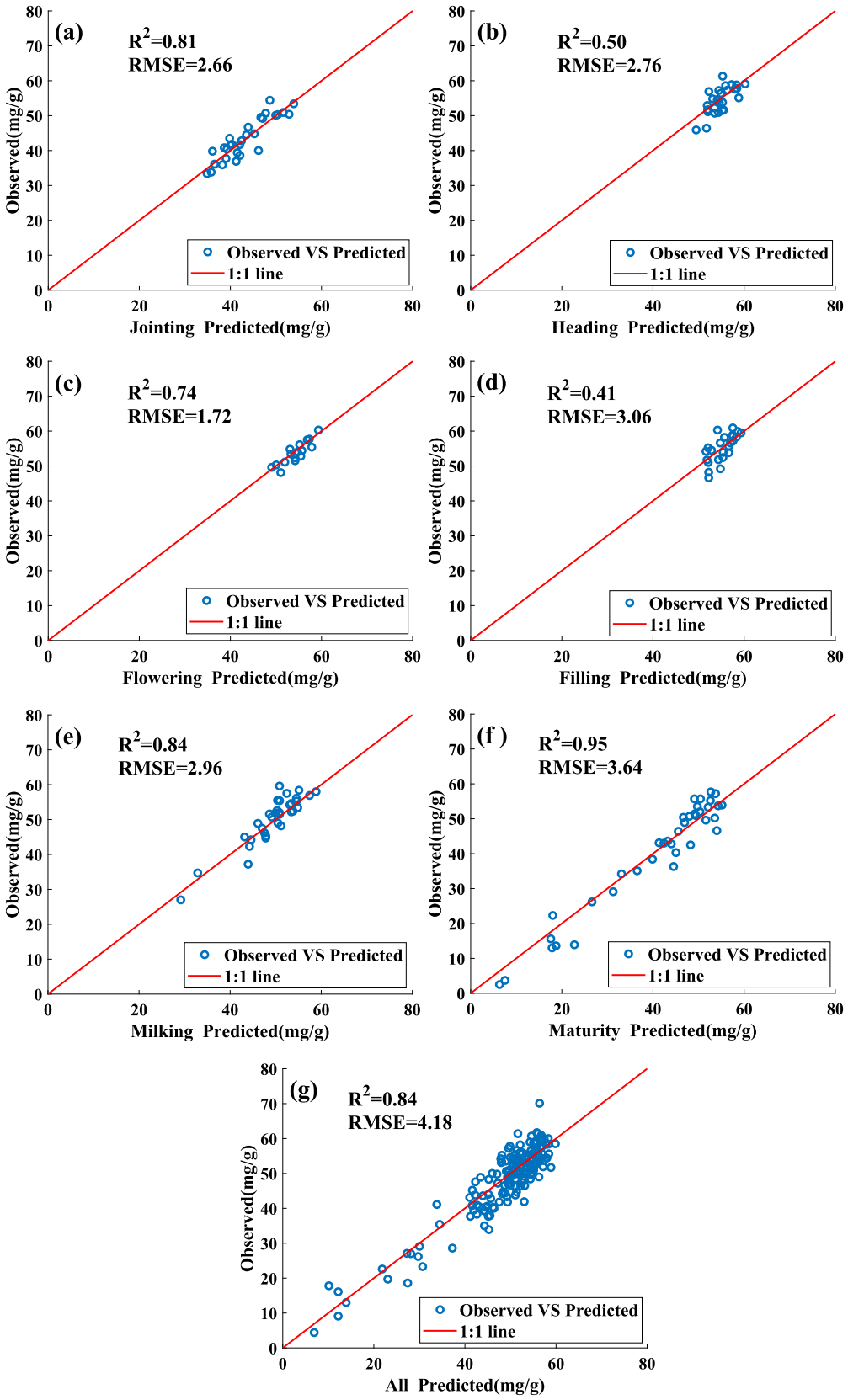


**Fig. 9.** The prediction accuracy of chlorophyll content of winter wheat based on Gradient Boosted Regression Trees model for seven growth periods. The prediction results of the training set of the seven growth periods are respectively (a) (b) (c) (d) (e) (f) (g). Among them (a) is the jointing period, (b) is the heading period, (c) is the flowering period, (d) is the filling period, (e) is the milk maturity period, (f) is the maturity period, and (g) is the whole grow period.

**Table 3**
Comparison of the prediction results of the Ridge regression model and the Gradient Boosted Regression Trees model in the seven growth periods, including the model determination coefficient ($R^2$) and the root mean square error (RMSE).

| Growth stage | $R^2_{Ridge}$ | $RMSE_{Ridge}$ | $R^2_{GBRT}$ | $RMSE_{GBRT}$ | $R^2_{BP}$ | $RMSE_{BP}$ |
|---|---|---|---|---|---|---|
| Jointing | 0.73 | 4.05 | **0.81** | **2.66** | 0.75 | 3.85 |
| Heading | 0.45 | 2.99 | **0.50** | **2.76** | 0.46 | 2.81 |
| Flowering | 0.67 | 1.95 | **0.74** | **1.72** | 0.67 | 1.56 |
| Filling | 0.29 | 3.37 | **0.41** | **3.06** | 0.32 | 2.34 |
| Milking | 0.70 | 4.13 | **0.84** | **2.96** | 0.70 | 3.76 |
| Maturity | 0.89 | 3.89 | **0.95** | **3.64** | 0.91 | 4.91 |
| Whole | 0.83 | 3.93 | **0.84** | **4.18** | 0.84 | 4.29 |

GBRT model and the EN-GBRT model is only 1%. After a comprehensive comparison, the GBRT model is more effective than the EN-GBRT model in retrieving leaf chlorophyll content ($R^2_{GBRT} \geq R^2_{EN\text{-}GBRT}$) in the six different growth stages of winter wheat. But for the whole growth period, the effect of the EN-GBRT prediction model is better than that of GBRT ($R^2_{EN\text{-}GBRT} > R^2_{GBRT}$, $RMSE_{EN\text{-}GBRT} < RMSE_{GBRT}$).

In conclusion, the optimal prediction model in this paper is: in the six different growth periods of winter wheat, from jointing to maturity, the GBRT prediction model has the best effect when inverting the chlorophyll content. The EN-GBRT prediction model is better when inverting the chlorophyll content during the whole growth period.

## 5. Discussion

### *5.1. The challenges of monitoring leaf-level physiological traits via canopy-level hyperspectral data*

Although SPAD value has a high consistency with chlorophyll content in leaves, it depends on the measurement of transmitted light in leaves under ideal conditions. Therefore, estimating chlorophyll content at leaf level using canopy level hyperspectral data is challenging.

Milas et al. (2018) believe that due to the complexity of the decoupled spectral signal, chlorophyll content and leaf area index are used to construct canopy chlorophyll content (LCC*LAI) for experimental analysis. As Knyazikhin et al. (2013) suggested that the contribution of crop leaf and canopy reflectance to monitoring leaf chlorophyll content should be separated. Therefore, they do not recommend using canopy spectral reflectance to directly estimate leaf chlorophyll content.

However, Daughtry et al. (2000) study showed that using canopy spectral vegetation index pair can estimate leaf chlorophyll concentration with minimal LAI and background reflectance interference. And some studies have shown that the SPAD value of the functional leaf (the top second leaf) has a strong correlation with canopy spectrum. Zhao et al. (2014) studied the correlation between SPAD value of functional leaves of winter wheat and canopy hyperspectral reflectance data, and established a quantitative relationship between chlorophyll content and canopy reflected light spectrum, which is of great significance for rapid and non-destructive detection of plant chlorophyll content. Vincini et al. (2014) used canopy hyperspectral reflectance data and leaf chlorophyll content to compare the sensitivity of different vegetation indices (VIs) to leaf chlorophyll content at canopy scale under different spectral resolutions. Li et al. (2015) used canopy hyperspectral reflectance to simulate the spectral reflectance of GF-1, and constructed the remote sensing monitoring model of relative chlorophyll content (SPAD) of winter wheat leaves. The studies of Dian et al. (2016) and Blackburn (1998) showed that crop hyperspectral data generated at leaf and canopy levels could significantly affect leaf chlorophyll content at different band positions and bandwidths.

To sum up, more influencing factors need to be considered to monitor leaf chlorophyll content using crop canopy height spectral data.

Crop canopy spectrum is composed of crop spectral characteristics and background soil characteristics. There are many factors that affect crop canopy spectrum, such as sunlight incidence Angle, bidirectional reflection, aerosol, wind speed and many other external factors in the process of spectral determination. When using hyperspectral remote sensing technology to monitor SPAD in winter wheat leaves, the influence of canopy structure on chlorophyll content estimation should be minimized. Jay et al. (2017) have summarized some effects of canopy structure minimization on estimation of chlorophyll content in leaves. So, our future research should focus on proposing more effective methods to solve this problem.

### *5.2. Analysis of model prediction results*

The changes in the chlorophyll content of the six different growth stages of winter wheat in this paper are shown in Fig. 12. The chlorophyll content prediction model was established for the seven growth periods (including the whole growth period) of winter wheat. Including Ridge, BP, GBRT, EN-GBRT, EN-BP five prediction models.

The ridge regression prediction model (Tibshirani and Taylor, 2012) reduces the standard error of the model by adding the regularized L2 norm as a penalty term in the multiple linear regression model. A multiple linear regression with hyperspectral full-band data as the independent variable was established to realize the prediction of the chlorophyll content of winter wheat leaves. However, the interference of multicollinearity exists in the multiple linear regression model, so it will cause the prediction result to be distorted and reduce the accuracy. Therefore, the nonlinear method is introduced in this paper to establish the prediction model. The gradient boosting regression tree model builds a more powerful model by merging multiple regression trees (Friedman, 2002). It uses a continuous way to construct trees, and each tree is trying to correct the error of the previous tree. Such a model occupies less memory and has a faster prediction speed. However, this model has a disadvantage that it is extremely sensitive to parameter settings. If the parameters are set incorrectly, the accuracy of the model will be very low. Therefore, the parameter selection of this model is very important (Marcoulides, 2004). The prediction of different growth periods requires re-adjustment of parameters due to different samples. Studies have shown that too many iterations will not reduce the performance of the GBRT model on the test set (Nauss and Kokhanovsky, 2006). For the number of splits of the decision tree, each split is actually a rectangular space of the feature variable, corresponding to a conditional selection process. In many cases, as long as the number of iterations is large enough, it is sufficient to use a decision tree or even a tree stump with a small number of splits to achieve the best results.

Compare the Ridge, BP, GBRT models, and the results are shown in Fig. 13. Compared with the BP neural network model, the prediction error of the GBDT model is small and the prediction accuracy is high. The GBDT model can obtain more accurate prediction results, and can dig deeper into the hidden nonlinear relationship between the variable and the predicted chlorophyll content (Cheng et al., 2019). From the above analysis, we can see that the spectral characteristics of the winter wheat leaf chlorophyll content prediction model tends to be more nonlinear, so the nonlinear model is better than the linear model. At the same time, this is also because of the interference of multicollinearity between the independent variable data, which leads to the low prediction ability of the Ridge model, so the nonlinear model is better than the linear model.

Since there are many bands of hyperspectral data, there will be multicollinearity interference and noise influence among the data. Therefore, this paper introduces the EN dimension reduction algorithm to remove the redundant data in the spectrum, retain all spectral information to the maximum extent, and obtain the data set with the minimum redundancy. Compared with the BP model, the prediction effect of the EN-BP model is better. Because of the multi-collinearity

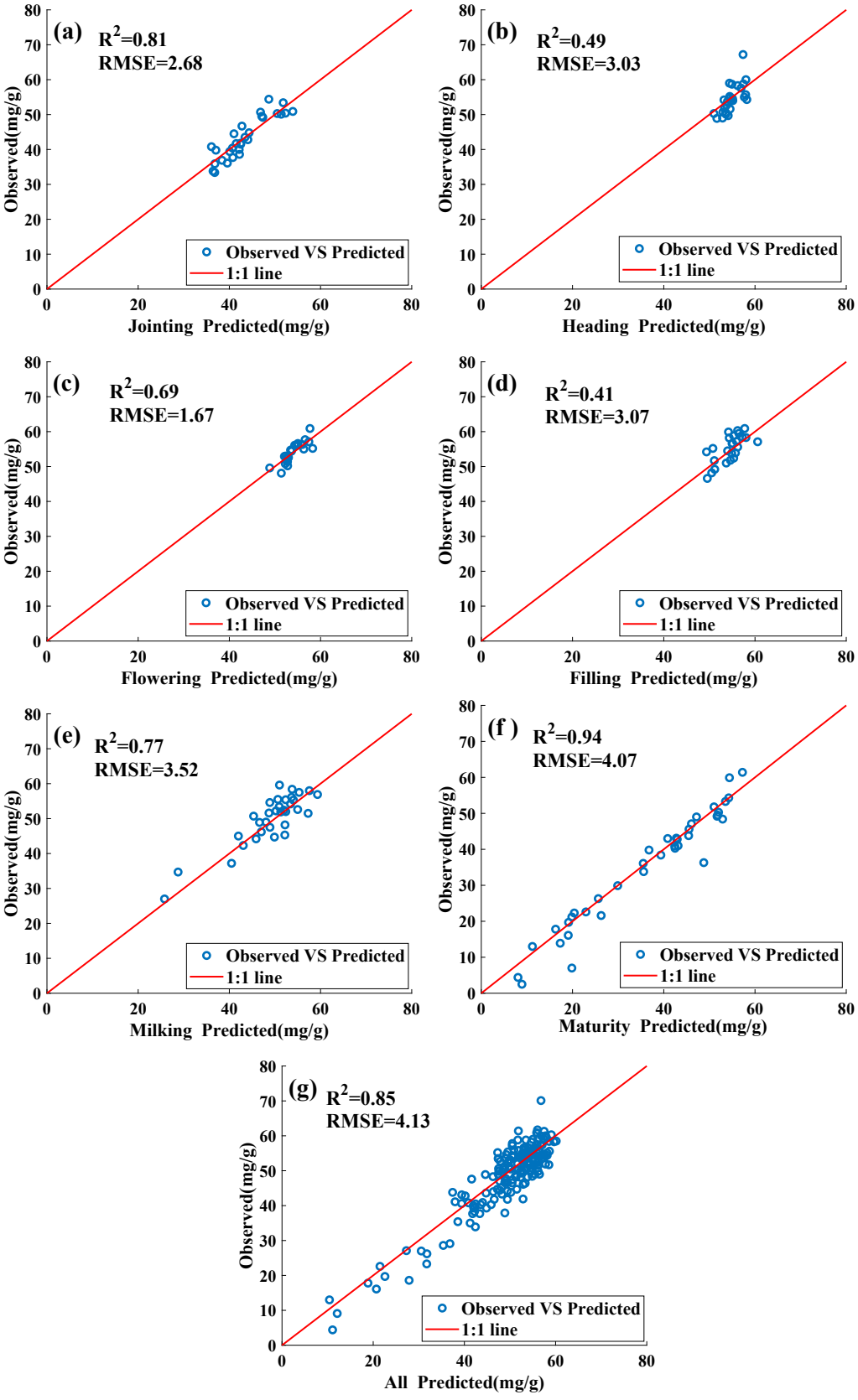


**Fig. 10.** The prediction accuracy of chlorophyll content of winter wheat based on EN-GBRT model for seven growth periods. The prediction results of the training set of the seven growth periods are respectively (a) (b) (c) (d) (e) (f) (g). Among them (a) is the jointing period, (b) is the heading period, (c) is the flowering period, (d) is the filling period, (e) is the milk maturity period, (f) is the maturity period, and (g) is the whole grow period. EN-GBRT: The gradient boosting regression tree model based on the elastic net dimensionality reduction algorithm.

**Table 4**
Comparison of the prediction results of the gradient boosting regression tree (GBRT) model and the gradient boosting regression tree model based on the elastic net dimensionality reduction algorithm (EN-GBRT) model in the seven growth periods.

| Growth stage | $R^2_{GBRT}$ | $RMSE_{GBRT}$ | $R^2_{EN\text{-}GBRT}$ | $RMSE_{EN\text{-}GBRT}$ |
|---|---|---|---|---|
| Jointing | **0.81** | **2.66** | 0.81 | 2.68 |
| Heading | **0.50** | **2.76** | 0.49 | 3.03 |
| Flowering | **0.74** | **1.72** | 0.69 | 1.67 |
| Filling | **0.41** | **3.06** | 0.41 | 3.07 |
| Milking | **0.84** | **2.96** | 0.77 | 3.52 |
| Maturity | **0.95** | **3.64** | 0.94 | 4.07 |
| Whole | 0.84 | 4.18 | **0.85** | **4.13** |

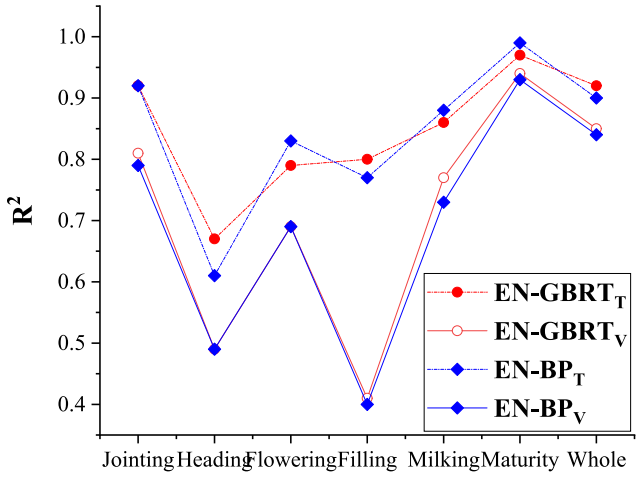


**Fig. 11.** The training and verification results of EN-GBRT and EN-BP models in seven growth stages are compared. EN-GBRT: the gradient boosting regression tree model based on elastic net dimensionality reduction; EN-BP: the feedforward backward neural network model based on elastic net dimensionality reduction.

**Table 5**
Detailed design information of EN-BP neural network estimation model for different growth stages.

| Growth stage | Maximum training times | Proportion of dropout | Number of neurons | $R^2_{EN\text{-}BP}$ | $RMSE_{EN\text{-}BP}$ |
|---|---|---|---|---|---|
| Jointing | 2000 | 0.34 | 18:5:1 | 0.79 | 3.89 |
| Heading | 2500 | 0.45 | 19:5:1 | 0.49 | 1.99 |
| Flowering | 3800 | 0.75 | 24:5:1 | 0.69 | 1.46 |
| Filling | 3000 | 0.75 | 15:10:1 | 0.40 | 3.16 |
| Milking | 5000 | 0.55 | 24:5:1 | 0.73 | 5.92 |
| Maturity | 8000 | 0.35 | 24:5:1 | 0.93 | 3.88 |
| Whole | 8000 | 0.35 | 20:5:1 | 0.84 | 3.81 |

between the input neurons of the BP network, the generalization ability of the network is not high, and the prediction performance of the network is reduced (Wang et al., 2005, Xiu-Zhi et al., 2013). Therefore, the EN-BP model after dimensionality reduction has a better effect. Elastic net dimensionality reduction algorithm can effectively remove the redundancy between hyperspectral data. Therefore, after removing the redundancy and then fusing the model established by the BP neural network, a better prediction effect can be obtained.

Finally, this paper compares the prediction results of the GBRT model and the EN-GBRT model. As shown in Fig. 14, the prediction results of the GBRT model are more accurate and more effective than the EN-GBRT model in six different growth periods. However, if only the chlorophyll content of winter wheat during the whole growth period is retrieved, the EN-GBRT model is more effective. The prediction

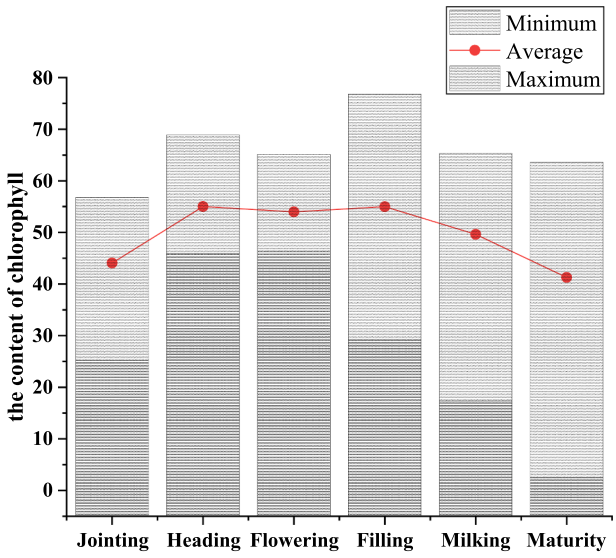


**Fig. 12.** The trends of maximum, minimum and average chlorophyll content in winter wheat at six growth stages. The six growth stages were the jointing stage, heading stage, flowering stage, grain-filling stage, milking stage and the maturity stage.

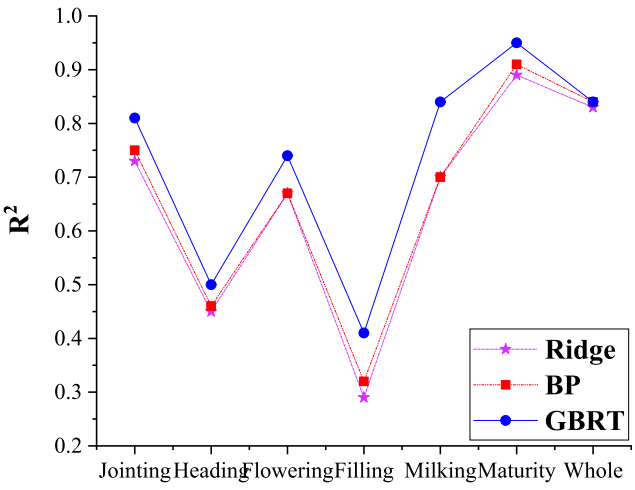


**Fig. 13.** Comparison of the model determination coefficients ($R^2$) between the Ridge regression linear prediction model, the Back Propagation Neural Network (BP) nonlinear prediction model and the Gradient Boosting Regression Tree (GBRT) nonlinear prediction model.

accuracy of GBRT model and EN-GBRT model in different growth periods is not much different, indicating that the GBRT model is not sensitive to multicollinearity between independent variables. In the process of tuning the GBRT model, the "max_feature" parameter represents the selection of $\sqrt{n}$ features in the data set for modeling each time when "sqrt" is selected. This parameter can automatically remove the interference of multicollinearity between sample features, thereby reducing data redundancy and achieving the purpose of improving model accuracy (Xiao-Ting and Yan, 2009). The GBRT model is more suitable for establishing a prediction model of chlorophyll content based on hyperspectral full-band data in different growth periods.

In summary, the best prediction model in this paper is: if the chlorophyll content of winter wheat is retrieved in six different growth periods, that is, from the jointing stage to the maturity stage, the GBRT prediction model is selected. Due to the large amount of data throughout the growth period and the increase in redundancy between data, the elastic net dimensionality reduction algorithm greatly reduces the

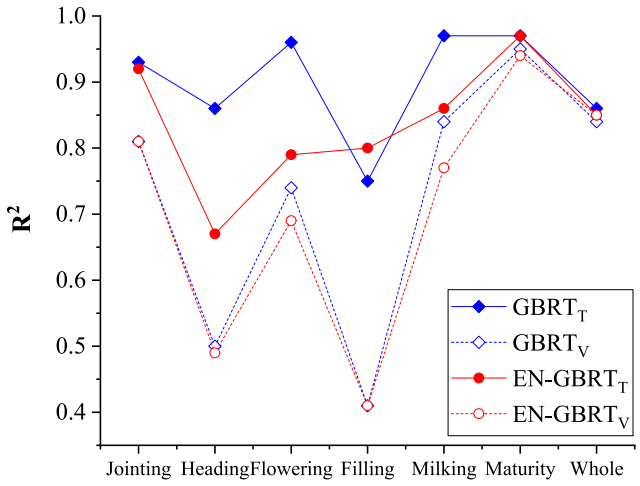


**Fig. 14.** Comparison of the model determination coefficients ($R^2$) between the training set and the validation set of the GBRT model and the EN-GBRT model. GBRT: Gradient Boosting Regression Tree, EN-GBRT: the gradient boosting regression tree model based on elastic net dimensionality reduction.

redundancy between data and effectively retains effective spectral information. Therefore, the winter wheat content prediction model established by the fusion of the GBRT model has a better effect at this time. If you only invert the entire growth period data set, it is best to choose the EN-GBRT prediction model.

### 5.3. Analysis of model universality

This paper is based on spectral information to retrieve crop physiological parameters, which is of great significance for ensuring food and environmental safety. This study provides techniques and methods for timely and accurate monitoring and diagnosis of winter wheat chlorophyll content and obtaining winter wheat growth information. This study fully improved the utilization of spectral data and increases the versatility of the model. So as to guide the field management of winter wheat and promote the development of precision agriculture to provide data and technical support. This paper analyzed the various growth periods of winter wheat and perfected the problem of poor accuracy in retrieving different growth periods of a single model. Therefore, this paper has made a contribution to the establishment of the prediction model of winter wheat chlorophyll content. But this article still needs improvement, the analysis is as follows:

1. The research results of this paper were only carried out in a single year and under the same ecological environment, and there was a lack of systematic research on the differences caused by differences in different ecological environments. Moreover, different varieties of winter wheat may have different metabolic characteristics. Therefore, the conclusions of this study will to be further verified in different ecological conditions and different types of varieties.
2. The preprocessing method of hyperspectral data in this paper is relatively simple. Next, different preprocessing methods (such as log (1/R), R', (1/R)', etc.) can be performed on the spectral data to reduce the influence of background noise. At the same time, choose more dimensionality reduction methods to reduce the dimensionality of hyperspectral data, and then establish an empirical regression model, hoping to obtain a more accurate chlorophyll content estimation model.

Introduce more advanced dimensionality reduction methods, such as the geodesic-based sparse manifold hypergraph method (Duan et al., 2021). It mainly uses the neighborhood sparse representation model based on geodesic to reveal the sparse relationship between different manifold neighborhoods. With the help of sparse coefficients and geodesic distance, a pair of semi-supervised hypergraphs are constructed based on the geodesic sparse manifold hypergraphs to combine sample information during training and maintain the multivariate sparse manifold relationships in the embedding space. This method can discover more useful discriminant information, further improve the feature discrimination ability of the sample, and achieve the purpose of extracting useful discriminant features.

In order to effectively reduce the number of hyperspectral bands and retain some ideal internal information, the semi-supervised sparse manifold discriminant analysis method (Luo et al., 2016b) is also effective. This method can adaptively select data points from the same manifold based on the idea of manifold-based sparse representation. In addition, it constructs an intra-class graph, an inter-class graph, and an unsupervised graph with labeled and unlabeled samples to enhance inter-class separability and intra-class compactness. Therefore, it reveals the sparse manifold structure of the data to improve its ability to extract effective features.

(3) In the research process, only ground-measured samples were used, not combined with image data. Therefore, this research needs to further realize the prediction of chlorophyll content in different growth stages on the satellite image scale.

## 6. Conclusion

Ground hyperspectral data provides detailed vegetation reflectance information with thousands of bands, which leads to the dimensional disaster of regression fitting. Preserving effective spectral information while improving the accuracy of the prediction model is a challenging problem. In order to solve these problems, we propose to compare the linear regression model with the nonlinear regression model, and propose a nonlinear regression model coupled with elastic net dimensionality reduction algorithm. The research conclusions can be summarized as follows.

1. The linear regression algorithm takes the ridge regression algorithm as an example, and the nonlinear regression algorithm takes the BP neural network algorithm and the gradient boosting regression tree integration algorithm as an example to establish a model. The results show that for the full-band spectrum, the linear regression model is more affected due to the multicollinear interference in the data set, so the BP model and the GBRT model have better prediction effects than the Ridge model. After comparing the BP model with the GBRT model, the study found that the use of a single BP model is prone to defects such as low accuracy and poor generalization ability. This is because a single BP model will also be affected by the multicollinearity between input neurons, resulting in the model's prediction effect is not as good as the GBRT model. In contrast, for full-spectrum data, the GBRT model is established without assumptions and pre-presets, and can more effectively deal with multicollinearity and obtain more accurate and stable results. At the same time, it can model complex nonlinear data and capture the nonlinear effects of independent variables.
2. In this paper, a nonlinear regression model of coupled elastic net dimensionality reduction algorithm is established. The results show that the coupled EN-BP model is better than the single BP model. Because the collinearity between the data after dimensionality reduction is very low, that is, the data redundancy of the input neuron is reduced, so the prediction accuracy of the model is improved. Compared with the EN-GBRT model, the prediction effects of the GBRT model and the EN-GBRT model are similar. This is because the "max-features" parameter of the GBRT model can effectively filter features for modeling when selecting "sqrt" in the process of parameter adjustment, remove the interference of multicollinearity, and achieve the purpose of dimensionality reduction.

3. When predicting chlorophyll content in six different growth periods, the best prediction model selected is: GBRT model. In the jointing period ($R^2_{jointing} = 0.81$, RMSE = 2.66) and heading period ($R^2_{heading} = 0.50$, RMSE = 2.76), flowering period ($R^2_{flowering} = 0.74$, RMSE = 1.72), filling period ($R^2_{filling} = 0.41$, RMSE = 3.06), milking period ($R^2_{milking} = 0.84$, RMSE = 2.96) and maturity period ($R^2_{maturity} = 0.95$, RMSE = 3.64), the GBRT prediction model has the best effect. However, the effect of the EN-GBRT prediction model is also very good, with a smaller amount of data and less running time. When using all the data to invert the chlorophyll content of the whole growth period, the EN-GBRT prediction model is the best ($R^2_{whole} = 0.85$, RMSE = 4.13).

This paper has fully studied the BP neural network model and the gradient boosting regression tree model, which has important theoretical significance for the in-depth study of the use of machine learning models to establish vegetation physiological parameter estimation models. This study can provide a reference for making full use of hyperspectral data and establishing a fast, efficient, and non-destructive prediction model of winter wheat leaf chlorophyll content.

## CRediT authorship contribution statement

**Tianli Wang:** Methodology, Investigation, Writing – original draft, Writing – review & editing. **Maofang Gao:** Methodology, Writing – original draft, Writing – review & editing, Project administration. **Chunling Cao:** Writing – review & editing. **Jiong You:** Writing – review & editing. **Xiwang Zhang:** Writing – review & editing. **Lanzhi Shen:** Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was supported by the National Natural Science of China (grant number 41921001, 41871282), Independent research and development project of Academy of Agricultural Planning and Engineering, Ministry of Agriculture and Rural Affairs (grant number ZZYFXKFZ201904), Open Fund of Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education (Grant No. GTYR202005), and Research on key technologies of multi-satellite joint large area coverage imaging (grant number 20210201015G X).

## References

An, G., Xing, M., He, B., Liao, C., Kang, H., 2020. Using machine learning for estimating rice chlorophyll content from in situ hyperspectral data. Remote Sens. 12, 3104.
Atzberger, C., Guérif, M., Baret, F., Werner, W., 2010. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. Comput. Electron. Agric. 73, 165–173.
Blackburn, 1998. Quantifying chlorophylls and caroteniods at leaf and canopy scales: an evaluation of some hyperspectral approaches. Remote Sens. Environ. 66, 273–285.
Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159.
Chen, L., Huang, J.F., Wang, F.M., Tang, Y.L., 2007. Comparison between back propagation neural network and regression models for the estimation of pigment content in rice leaves and panicles using hyperspectral data. Int. J. Remote Sens. 28, 3457–3478.
Cheng, J., Li, G., Chen, X.H., 2019. Research on travel time prediction model of freeway based on gradient boosting decision tree. IEEE Access 7, 7466–7480.
Chunjiang, Z., Wenjiang, H., Jihua, W., Liangyun, L., Xiaoyu, S., Zhihong, M., Cunjun, L., 2006. Extracting winter wheat chlorophyll concentration vertical distribution based on bidirectional canopy reflected spectrum. Trans. Chinese Soc. Agric. Eng. 22, 104–109.
Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtrey, J.E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sens. Environ. 74, 229–239.
Dian, Y.Y., Le, Y., Fang, S.H., Xu, Y.R., Yao, C.H., Liu, G., 2016. Influence of spectral bandwidth and position on chlorophyll content retrieval at leaf and canopy levels. J. Indian Soc. Remote Sens. 44, 583–593.
Duan, Y., Huang, H., Wang, T., 2021. Semisupervised feature extraction of hyperspectral image using nonlinear geodesic sparse hypergraphs. IEEE Trans. Geosci. Remote Sens.
Feng, W., Yao, X., Tian, Y., Cao, W., Zhu, Y., 2008. Monitoring leaf pigment status with hyperspectral remote sensing in wheat. Aust. J. Agric. Res. 59, 748–760.
Friedman, 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.
Friedman, J.H., 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis, 38, 367-378.
Fushiki, 2011. Estimation of prediction error by using K-fold cross-validation. Statist. Comput. 21, 137–146.
Guo, L., Qiao, L., Chengcai, L.I., Zhou, Y., 2016. Weather characteristics of constant temperature and its predictability analysis in Langfang City. Journal of Arid Meteorology 1, 195–201.
Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: applications to nonorthogonal problems. Technometrics 12, 69–72.
Huang, H.L., Niu, B., 2011. Comparison of ridge parameter determination. Sci. Survey. Map. 36, 31–32.
Huang, J.R., Sun, J.Y., Liao, H.J., Liu, X.D., 2015. Detection of brown planthopper infestation based on SPAD and spectral data from rice under different rates of nitrogen fertilizer. Precis. Agric. 16, 148–163.
Jay, S., Gorretta, N., Morel, J., Maupas, F., Bendoula, R., Rabatel, G., Dutartre, D., Comar, A., Baret, F., 2017. Estimating leaf chlorophyll content in sugar beet canopies using millimeter- to centimeter-scale reflectance imagery. Remote Sens. Environ. 198, 173–186.
Jiang, J.B., Chen, Y.H., Huang, W.J., 2010. Using hyperspectral remote sensing to estimate canopy chlorophyll density of wheat under yellow rust stress. Spectroscopy Spectral Anal. 30, 2243–2247.
Knyazikhin, Y., Schull, M.A., Stenberg, P., Mottus, M., Rautiainen, M., Yang, Y., Marshak, A., Carmona, P.L., Kaufmann, R.K., Lewis, P., Disney, M.I., Vanderbilt, V., Davis, A.B., Baret, F., Jacquemoud, S., Lyapustin, A., Myneni, R.B., 2013. Hyperspectral remote sensing of foliar nitrogen content. PNAS 110, E185–E192.
Kong, W., Huang, W., Zhou, X., Ye, H., Dong, Y., Raffaele, C., 2017. Off-nadir hyperspectral sensing for estimation of vertical profile of leaf chlorophyll content within wheat canopies. Sensors 17, 2711.
Li, C., Liu, W., Ding, X., Zou, J., Ma, J., Wang, F., Lin, Z., 2017. Inversion model for chlorophyll content of Phragmites under different sample areas based on red edge parameter. Ecol. Sci. 36, 66–73.
Li, F., Wang, L., Liu, J., Chang, Q., 2015. Remote sensing estimation of SPAD value for wheat leaf based on GF-1 data. Trans. Chinese Soc. Agric. Mach. 46, 273–281.
Liu, J., Sui, C., Deng, D., Wang, J., Feng, B., Liu, W., Wu, C., 2016. Representing conditional preference by boosted regression trees for recommendation. Inf. Sci. 327, 1–20.
Luo, D., Chang, Q., Yanbing, Q.I., Yuanyuan, L.I., Song, L.I., 2016a. Estimation model for chlorophyll content in winter wheat canopy based on spectral indices. J. Triticeae Crops 36, 1225–1233.
Luo, F., Huang, H., Ma, Z., Liu, J., 2016b. Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images. IEEE Trans. Geosci. Remote Sens. 54, 6197–6210.
Macedo, p., 2017. Ridge regression and generalized maximum entropy: an improved version of the ridge–GME parameter estimator. Commun. Statist.- Simulat. Computat. 46, 3527–3539.
Marcoulides, g. a., 2004. The elements of statistical learning: Data mining, inference and prediction. Struct. Equat. Model. – A Multidiscipl. J. 11, 150–151.
Milas, A.S., Romanko, M., Reil, P., Abeysinghe, T., Marambe, A., 2018. The importance of leaf area index in mapping chlorophyll content of corn under different agricultural treatments using UAV images. Int. J. Remote Sens. 39, 5415–5431.
Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front. Neurorob. 7, 21.
Nauss, T., Kokhanovsky, A.A., 2006. Discriminating raining from non-raining clouds at mid-latitudes using multispectral satellite data. Atmos. Chem. Phys. 6, 5031–5036.
Pham, T.D., Le, N.N., Ha, N.T., Nguyen, L.V., Xia, J., Yokoya, N., To, T.T., Trinh, H.X., Kieu, L.Q., Takeuchi, W., 2020. Estimating mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused sentinel-2 and ALOS-2 PALSAR-2 data in can gio biosphere reserve, Vietnam. Remote Sens. 12 (5), 777. https://doi.org/10.3390/rs12050777.
Qian, B., Huang, W., Huichun, Y., 2021. Inversion of winter wheat chlorophyll contents based on improved algorithms for red edge position. Nongye Gongcheng Xuebao/ Trans. Chinese Soc. Agric. Eng. 36, 162–170.
Qu, Y.H., Wang, J.D., Wan, H.W., Li, X.W., Zhou, G.Q., 2008. A Bayesian network algorithm for retrieving the characterization of land surface vegetation. Remote Sens. Environ. 112, 613–622.
Hecht-Nielsen, R., 1988. Theory of the backpropagation neural network. Neural Networks 1, 445. https://doi.org/10.1016/0893-6080(88)90469-8.
Sawuti, R., Abula, A., Kasimu, N., Hu, L.I., Aishanjiang, W., Kahaer, Y., Xiaohang, L.I., 2019. Spectral estimation of chlorophyll content in spring wheat leaves based on fractional differential. J. Triticeae Crops 39, 738–746.
Schlichting, A.F., Bonfim-Silva, E.M., Silva, M.D.C., Pietro-Souza, W., Silva, T., Farias, L. D.N., 2015. Efficiency of portable chlorophyll meters in assessing the nutritional status of wheat plants. Rev.bras.eng.agríc.ambient 19, 1148–1151.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learni. Res. 15, 1929–1958.

Srivastava, P.K., Gupta, M., Singh, U., Prasad, R., Pandey, P.C., Raghubanshi, A.S., Petropoulos, G.P., 2021. Sensitivity analysis of artificial neural network for chlorophyll prediction using hyperspectral data. Environ. Dev. Sustain. 23, 5504–5519.

Sun, B., Chang, Q., Liu, M., 2017. Inversion chlorophyll mass fraction in winter wheat canopy by hyperspectral reflectance. Acta Agriculturae Boreali-occidentalis Sinica 26, 552–559.

Tibshirani, R.J., Taylor, J., 2012. Degrees of freedom in lasso problems. Ann. Stat. 40, 1198–1232.

Vincini, M., Amaducci, S., Frazzi, E., 2014. Empirical estimation of leaf chlorophyll density in winter wheat canopies using sentinel-2 spectral resolution. IEEE Trans. Geosci. Remote Sens. 52, 3220–3235.

Wang, Z.Z., Jin, J.L., Zheng, Z.S., Zhang, L.L., 2005. Improved BP neural network and its application in forecasting groundwater level in Jinan city. Hydro-Sci. Eng. 71–74.

Xiao-Ting, W.U., Yan, D.Q., 2009. Analysis and research on method of data dimensionality reduction. Appl. Res. Comput. 26, 2832–2835.

Xiu-Zhi, S., Yong-Meng, W.U., Li-Zhong, T., Xuan-Dong, H., 2013. Application of neural network model with partial least-square regression in prediction of peak velocity of blasting vibration. J. Vibrat. Shock 32, 45–49.

Yao, Y.-X., Yun, Z., Liu, X.Y., Nan, H., 2014. The Research and Application of Remote Sensing Technology in MaizeYield Estimation of Jilin. IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA), 2014, Sep 29-30 Ottawa, CANADA. 1003-1005.

Yu, Q., Yang, G., Wang, C., 2019. Chlorophyll inversion of winter wheat based on ground hyperspectral data and PROSAIL model. Sci. Surv. Map. 44, 96.

Zhang, L., Zhang, Z., Luo, Y., Cao, J., Tao, F., 2019. Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in china using machine learning approaches. Remote Sens. 12 (1), 21. https://doi.org/10.3390/rs12010021.

Zhao, J., Feng, M., Chao, W., Yang, W., Li, Z., Zhu, Z., Peng, R., Liu, T., Wang, H., 2014. Simulating the content of chlorophyll in winter wheat based on spectral vegetation index. J. Shanxi Agricul. University(Natural Science Edition) 34, 391–396.

Zhu, X.K., Sheng, H.J., Gu, J., 2005. Primary study on application of SPAD value to estimate chlorophyll and nitrogen content in wheat leaves. Acta Tritical. Crops 25, 46–50.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. Roy. Statist. Soc. Ser. B-Statist. Methodol. 67, 301–320.