https://doi.org/10.17221/421/2023-PSE

Hyperspectral analysis of the content of the alkali-hydrolysed nitrogen in the soil of a millet field

Tingyu Zhu[1], Zhiqiang Wang[1], Zilin Zhang[1], Xiuhan He[1], Gangao Li[1], Zongbao Huang[2], Lili Guo[3], Zhiwei Li[1,2], Huiling Du[4]**

[1]*College of Agricultural Engineering, Shanxi Agricultural University, Jinzhong, P.R. China*
[2]*College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong, P.R. China*
[3]*Agricultural Extension Center, Qin County, Changzhi, P.R. China*
[4]*Department of Basic Sciences, Shanxi Agricultural University, Jinzhong, P.R. China*
**Corresponding authors: lizhiweitong@163.com; duhuiling66@163.com*

Citation: Zhu T.Y., Wang Z.Q., Zhang Z.L., He X.H., Li G.A., Huang Z.B., Guo L.L., Li Z.W., Du H.L. (2023): Hyperspectral analysis of the content of the alkali-hydrolysed nitrogen in the soil of a millet field. Plant Soil Environ., 69: 596–607.

Abstract: Hyperspectral imaging technology has emerged as a prominent research area for quantitatively estimating soil nutrient content owing to its non-destructive, rapid, and convenient features. Our work collected the data from soil samples using the hyperspectrometer. Then, the data were processed. The competitive adaptive reweighted sampling (CARS) algorithm reduced the original 148 bands to 13, which accounted for 8.8% of the total bands. These selected bands possess a certain level of interpretability. Based on the modelling results, it can be concluded that the prediction model constructed by the least squares support vector machine (LSSVM) exhibited the highest accuracy. The coefficient determination, root mean square error, and ratio performance deviation were 0.8295, 2.95, and 2.42, respectively. These findings can provide theoretical support for the application of hyperspectral technology in detecting the content of the AHN in soil. Moreover, they can also serve as a reference for the rapid detection of other soil components.

Keywords: modern precision agriculture; soil fertility; characteristic bands selection; model building; content prediction

Nitrogen, one of the key nutrients essential for plant growth in soil, plays a crucial role. Organic matter in the soil can be decomposed by microorganisms into inorganic nitrogen. The level of alkali-hydrolysed nitrogen (AHN) is increased, which is beneficial for plant absorption and utilisation. As a major source of nutrients for plant growth, soil nutrient content is an important indicator for evaluating the environment for plant growth. Rapid access to soil nutrient information is an important requirement for the development of modern precision agriculture. As a major component of soil, soil nitrogen is an important indicator of soil fertility. Under certain conditions, the nitrogen content level marks the soil fertility level. It is the main source of nutrients needed for plant growth. Soil nitrogen plays a critical role in plant growth and development. Nitrogen deficiency affects the protein content of crops. When organic matter decomposes, most of the nitrogen undergoes conversion to inorganic nitrogen and is then released into the soil (Tahmasbian et al. 2018). If nitrogen in the soil is not released and converted

Supported by the Major Special Projects of National Key R&D, Project No. 2021YFD1600301-4; by the Major Special Projects of National Key R&D, Project No. 2017YFD0701501; by the Major Special Projects of Shanxi Province Key R&D, Project No. 201903D211005; by the Central Government Guides Local Funds for Scientific and Technological Development, Project No. YDZJSX20231A042; by the Construction Project of Shanxi Modern Agricultural Industry Technology System; and by the Major Projects of Shanxi Province Key R&D, Project No. 2022ZDYF119.

https://doi.org/10.17221/421/2023-PSE

in a timely manner, it may result in nitrogen loss and soil infertility. The decomposition of the alkali-hydrolysed nitrogen (AHN) can facilitate the transformation of organic nitrogen and sustain soil fertility. Therefore, monitoring the content of the AHN in soil has become a crucial task. Given the significance of establishing AHN levels for advancing soil formula fertilisation and precision agriculture, researchers in relevant fields have dedicated themselves to exploring expedited and real-time methods for assessing AHN content. However, the conventional method of determining the AHN content involves laboratory analysis through chemical determination. The results obtained are highly accurate. However, the testing procedure is demanding, lengthy, and expensive in terms of both time and money. Moreover, the laboratory analysis process requires a series of chemical pretreatments for samples, which inevitably generate harmful chemical waste to the environment. The reasons mentioned above render laboratory chemical determination methods inadequate for fulfilling the needs of monitoring the content of the AHN in large-scale, rapid and dynamic scenarios.

Hyperspectral imaging technology is a research focus for scientists in numerous agriculture-related institutes. This has been widely used in soil classification, composition detection and other similar research. At present, visible-near infrared (VNIR) hyperspectral technology is the primary means for rapidly detecting the AHN content (Pechanec et al. 2021). This technology has increasingly become a valuable method for acquiring soil properties due to its advantages of speed, simplicity, non-contact and non-destructiveness. In recent years, extensive research has focused on collecting VNIR spectra in laboratory conditions and utilising them to predict soil property contents. A well-established soil sample pretreatment system has also been developed, including sample drying, grinding, and screening. This process significantly decreases the impact of environmental disruptions like soil texture, surface particles, or moisture on spectral measurements.

Soil, a product of rock weathering, comprises complex aggregate. The spectral characteristics of soil are primarily influenced by the molecular vibrations and electronic transitions of various components within the soil, which can generate distinct peaks in different bands. Due to variations in soil components, the electromagnetic radiation absorbed by each component differs from that of the external light source, resulting in different jump energy level differences. Consequently, distinct absorption characteristic bands are present in the soil spectral curve. This forms the basis for spectral prediction modelling of soil constituents. The soil structure is complex, and its spectrum results from various components' comprehensive superposition of reflectance. The spectral information is frequently intertwined with the feedback information of various factors, such as salt and water. This information could cause non-specific soil spectra, thereby impacting the accuracy of spectral data. In addition, modelling accuracy using raw spectral data is frequently low owing to the interference from the instrument's own noise signal, spectral scattering and spectral covariance. Therefore, removing inaccurate information from spectral data and improving model accuracy have been topics of concern among scholars in recent years. Numerous studies have demonstrated that employing spectral smooth denoising, mathematical transformation, and other preprocessing methods can substantially enhance the dependability and precision of the spectral data. It has been found that utilising suitable regression modelling algorithms can significantly enhance the predictive accuracy of models. Regression modelling algorithms typically comprise both linear and nonlinear regression models. Among the linear regression models commonly used are multiple linear regression, partial least squares regression and principal component regression analysis. Due to the intricate non-linear correlation between spectral reflectance information and soil properties, the nonlinear regression model may offer greater predictive power for soil property content. Therefore, researchers have applied various nonlinear models to analyse this area, such as artificial neural networks, support vector machine regression, random forest regression, and multivariate adaptive regression spline. These models have contributed to the research and analysis in this field. Liu and Xiao (2021) developed a predictive model for AHN in soil using ridge regression (RR) and partial least squares regression (PLSR). Chen et al. (2022) developed machine learning models based on the full and characteristic bands, respectively. The best prediction model for various nitrogen tests was obtained by thoroughly comparing the evaluation metrics. Zhang et al. (2022) developed a model utilising RR, PLSR, support vector machine regression (SVR), and random forest (RF) to estimate nitrogen content in winter wheat leaves quantitatively. Tang et al. (2022) investigated the optimal estimation

https://doi.org/10.17221/421/2023-PSE

model for soil total nitrogen using fractional order derivative (FOD) and found that this enhanced the model's estimation ability. Niu et al. (2023) analysed the spectral data with various pretreatment methods and concluded that the accuracy of the soil total nitrogen content prediction model was influenced by the chosen preprocessing method.

In this work, we examined the soil from the millet experiment field during various growth periods. Two hundred twenty-five soil samples were collected for analysis. We were using Visual Basic 6.0 (Microsoft, Redmond, USA) to build a hyperspectral data preprocessing program. Afterwards, we obtained the average hyperspectral data from 225 soil samples. Combining various preprocessing methods, the CARS algorithm was utilised for feature variable extraction of spectral data. Then, the prediction models of the AHN were established by combining three machine learning algorithms: PLSR, back propagation neural network (BPNN) and LSSVM. This work aims to establish the prediction model of the AHN content in soil to provide theoretical support for rapid detection.

MATERIAL AND METHODS

Experimental design and sample collection. This experiment was conducted at the experimental site. The region has a warm, temperate continental climate with four distinct seasons. The average annual temperature, average annual total rainfall, average annual sunshine hours, and average annual frost-free days are 10.4 °C, 397.1 mm, 2 527.5 h, and 179 days, respectively. The region is blessed with ample sunlight, a mild climate, and fertile soil. The main soil types in the area are loess-like calcareous brown soil and alluvial tidal soil, with a small amount of brown soil and sulfate salted tidal soil. This experiment utilised the "Jin Gu 21" sorghum cultivar. The fertiliser, purchased from Jingtaisifang (China) Biotechnology Co., Ltd, had an organic matter content of not less than 60%, an effective number of living bacteria of 300 million per gram, and a nitrogen content of not less than 6% in the fertiliser. All fertilisers are applied at once as a base fertiliser.

The test plot was 3 acres in size. The topography is flat, and there are differences in soil fertility. It is recommended to use the checkerboard sampling method, taking a mixed sample every 10 to 15 points (Niu et al. 2017, Li et al. 2022b). Before collecting the sample, it was necessary to scrape off the surface soil layer, which was 2 to 3 mm thick. Then, the soil auger was used to collect soil cores from 0–20 cm depth. Next, the soil samples taken at each sampling point were spread onto plastic sheets at the head of the field. Crush the large soil samples. After that, stones and plant and animal residues were removed. Finally, the soil was mixed thoroughly to make a composite soil sample. The soil samples are naturally air-dried in the laboratory. After grinding, the samples were sieved through a 250-micron mesh screen, and the sieved soil samples were individually packaged. For testing purposes, a total of 225 samples were collected from different soil fertility conditions during the seedling, jointing, tasseling, filling, and maturity stages, with 45 samples collected at each stage. The samples were recorded and numbered for future use. Using the quadrat method for sampling, each sample was divided into two parts. One portion was used for laboratory determination of AHN content, and the other portion was used to collect near-infrared hyperspectral images of the soil. The hyperspectral images of each sample were collected. Then, the physicochemical values were measured.

Hyperspectral image acquisition. The NIR hyperspectral images were acquired using the Starter Kit indoor mobile scanning platform (Headwall Photonics, Bolton, USA). The system consists of several components: an indoor mobile scanning platform, a micro-sized near-infrared hyperspectral imaging device with an aperture of 1.4 mm and a focal length of 25 mm, a light source, a controller, and a computer. The movement speed, push-sweep stroke, and distance between the lens and the surface of the soil sample for the system were set to 2.721 mm/s, 100 mm, and 28 cm, respectively, to obtain clear and undistorted images. This system captures 170 spectral bands from 900 to 1 700 nm with a spectral resolution of approximately 0.727 nm. The spectral range with a total of 148 bands, ranging from 950 to 1 650 nm, was chosen due to the significant reflectivity error near the measuring range.

First, fill the soil sample into a round vessel about 5 cm in diameter and 3 cm deep. Next, flatten and compact the soil. Finally, the circular vessel was placed on the mobile scanning platform to obtain its hyperspectral image.

To reduce the interference from the system light source and dark current and to calculate the relative reflectance spectral values of the scanned objects, black and white correction of the hyperspectral images was performed before the experiment to ensure

https://doi.org/10.17221/421/2023-PSE

the accuracy of the experimental results. During the calibration process, the first step was to close the lens cap of the hyperspectral camera and acquire a full black calibration image with a reflectance of 0. Next, remove the lens cap and scan the calibration whiteboard to capture an all-white calibration image with a reflectance greater than 99.9%. Finally, the hyperspectral image information was corrected by applying the black-white calibration formula. The formula is as follows:

$$R = \frac{R_0 - R_b}{R_w - R_b} \quad (1)$$

Where: R – corrected hyperspectral image; R_0 – original hyperspectral image; R_w – white background image with the standard white calibration plate (> 99.9% reflectance); R_b – dark background image with the lens cap closed (< 0% reflectance).

Physicochemical determination of the AHN. AHN was determined by the alkaline diffusion method using 1 mol/L NaOH (Qi et al. 2018, Xie et al. 2019). After passing the air-dried soil sample through the 250-micron sieve, 2 g of soil sample and 1 g of ferrous sulphate powder were weighed and distributed evenly in the outer chamber of the diffusion dish. Rotate the diffusion dish gently and horizontally to flatten the soil sample. Add 2 mL of a 2% boric acid solution and 1 titrimetric nitrogen mixing indicator (phenolphthalein indicator) in the inner chamber of the diffusion dish. Apply alkaline gel solution to the outer chamber edge of the diffusion dish. Cover the ground glass and rotate it several times so that the ground glass can be glued completely to the edge of the diffusion dish. Cross the two rubber bands in a crisscross pattern to tighten the loop. Next, rotate one side of the ground glass slowly to expose a slit in the outer chamber of the diffusion dish and add 10 mL of 1.2 mol/L NaOH solution quickly to it. Then, turn the ground glass tightly. The diffusion dish was rotated horizontally and gently on the laboratory bench to mix the solution well with the soil. It was then placed in an incubator at 40 °C. After 24 h, the ammonia absorbed in the boric acid solution in the inner chamber of the diffusion dish was determined by titration with 0.01 mol/L hydrochloric acid standard solution using a microburette. The endpoint was the colour change from blue to fuchsia. The content of the AHN was calculated according to Eq. 2. The content measured through the chemical method is shown in Table 1.

$$\text{AHN (mg/kg)} = \frac{\text{C} \times (\text{V} - \text{V}_0) \times 14 \times 1000}{\text{W}} \quad (2)$$

Where: AHN – alkali-hydrolysed nitrogen (mg/kg); C – concentration of the hydrochloric acid standard solution (mol/L); V – volume of the hydrochloric acid standard solution used in the sample determination (mL); V_0 – volume of the hydrochloric acid standard solution used in the blank determination (mL); 14 refers to the molecular weight of nitrogen (g/mol); 1 000 refers to the conversion factor; W – weight of the soil (g).

Extraction of spectral data. Hyperspectral images contain both spectral and image information of soil samples. Each pixel point on the image has a corresponding diffuse reflectance spectral curve. In order to facilitate the selection of regions of interest (ROI) and the batch extraction and processing of spectral data, the SpectralView software (Headwall Photonics, Bolton, USA) was used for data extraction in this work. The software had been developed using Visual Basic (version 6.0, Microsoft, Redmond, USA) for secondary development, which has pixel generation and batch processing capabilities. The elliptical model was used to determine the ROI centre coordinates, X/Y half-axis lengths (a, b), and X/Y axis spacing (Δx and Δy were set to 1). Following the principle of "left to right, top to bottom" in the target image, the pixels in the ROI were captured sequentially according to the Eq. 3, and the ROI coordinate matrix was generated. Images were imported using the SpectralView software, and reflectivity information was actively extracted based on the coordinate

Table 1. Characteristic statistics of the content (mg/kg) of alkali-hydrolysed nitrogen in the soil

Growth stages	Samples	Minimum	Maximum	Mean	Standard deviation
Seedling	45	31.65	100.84	51 ± 2.39	12.57
Jointing	45	28.79	74.05	59.06 ± 1.21	8.83
Tasseling	45	44.33	95.61	73.12 ± 1.75	10.58
Filling	45	54.39	174.42	81.67 ± 3.28	22.86
Maturity	45	24.42	92.01	41.76 ± 2.12	10.26
Total	225	24.42	174.42	61.32 ± 1.35	20.16

https://doi.org/10.17221/421/2023-PSE

matrix. The batch module could generate numerical calculations such as mean, mean deviation, and variance on demand. To extract as much as possible the whole region, we selected the ROI according to Figure 1. 28.8000 (96.000 × 3) pixels were extracted for each sample. According to Eq. 4, the arithmetic mean value was taken as the spectral curve of the sample, which was the basic data set for subsequent data processing.

$$\frac{x_1^2}{a^2}+\frac{y_1^2}{b^2}\le 1 \quad (3)$$

Where: a – long wheelbase departure of the ellipse; b – short wheelbase departure of the ellipse; x and y – coordinate parameters.

$$A_i=\frac{1}{n}\left(A_1+A_2+\ldots+A_n\right)=\frac{1}{n}\sum_{k=1}^{n}A_{i,k} \quad (4)$$

Where: A – arithmetic mean; n – number of pixels; i – number of samples.

Spectrum pretreatment. The structure of the soil is complex. Its spectrum is the effect of a comprehensive superposition of the reflectance of various components. The spectral information is often interspersed with feedback from many factors, such as salinity and moisture. This information can lead to non-specificity in soil spectra, affecting the spectral data's accuracy. In addition, the accuracy of predicting soil property content using raw spectral data is often low due to the interference of instrument noise, spectral scattering and spectral covariance. Numerous studies have shown that the preprocessing of spectral data can effectively suppress and eliminate instrumental noise in the spectral data and improve the performance of the established mode. Savitzky-Golay (S-G) has a good smoothing effect and can effectively remove high-frequency noise and interfering signals (Zimmermann and Kohler 2013, Zhang et al. 2021). Discrete wavelet transform (DWT) can separate high-frequency and low-frequency components, making the signal more distinct (Liu et al. 2021, Xu et al. 2021). Standard normal variate (SNV) allows the data to be unified with zero mean and unit variance across the feature space, eliminating possible scale differences between the data (Ye et al. 2020). Derivatives can show the slope, curvature, and other relevant characteristics of changes in the data, allowing a better understanding of the trends and dynamics of the data (Bhadra et al. 2020). The Mapminmax algorithm can map the raw data to a specified range (usually 0 to 1 or –1 to 1) to ensure that the data are in a uniform range during modelling and analysis, thus improving the accuracy and reliability of data processing (Sun et al. 2017). These algorithms provide functions such as noise reduction, feature extraction, data standardisation and scale unification, which can improve the accuracy and reliability of data processing. Therefore, preprocessing the effective spectra using the S-G, DWT, SNV, Derivative, and Mapminmax can reduce the noise before modelling and effectively improve the model's prediction accuracy. The results of the preprocessing are shown in Figure 2.

Extraction of feature band and model building. In spectral data analysis, spectrometers typically measure thousands of wavelength points. Some wavelengths may contain critical information relevant to the object of study, while others may contain less information or noise. By extracting the feature wavelengths, the dimensionality of the data can be reduced, and the efficiency of model training and prediction can be improved. The CARS algorithm is a widely used spectral variable selection technique in the field of spectral modelling (Xing et al. 2021, Li and Yang 2023). It is a feature variable selection method that combines Monte Carlo sampling (MCS) with PLS model regression coefficients, mimicking the "survival of the fittest" principle in Darwinian theory. In the CARS algorithm, points with greater absolute weight of regression coefficients in the PLS model are retained as the new subsets each time through adaptive reweighted sampling (ARS). Remove points with lower weights. Then, the PLS model was constructed using the new subsets. After multiple calculations, we selected the wavelengths with the smallest root mean squared error of cross-validation (RMSECV) in the PLS model as the characteristic wavelengths.

Table 2. Statistics of the alkali-hydrolysed nitrogen (AHN) in the soil

Pigment	Subset	Number of samples	Range	Mean	Standard deviation
			(mg/kg)		
AHN	calibration set	169	24.42–174.42	61.86 ± 2.59	21.25
	prediction set	56	28.71–87.63	59.69 ± 1.46	16.12

https://doi.org/10.17221/421/2023-PSE

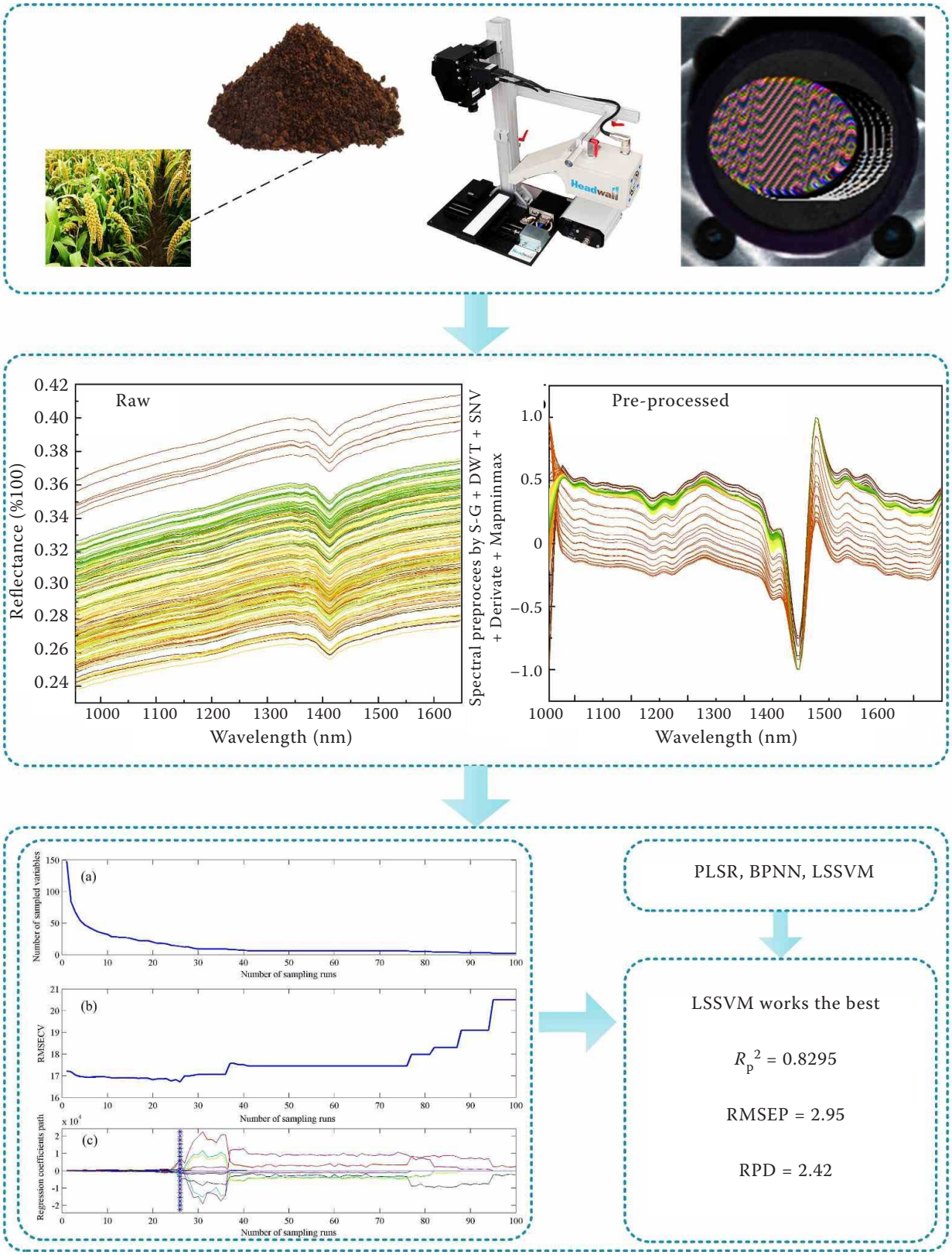

Figure 1. The schematic diagram of this work. RMSECV – root mean squared error of cross validation; PLSR – partial least squares regression; BPNN – back propagation neural network; LSSVM – least squares support vector machine; R_p^2 – determination coefficient of the prediction set; RMSEP – root mean square error of the prediction set; RPD – ratio performance deviation

https://doi.org/10.17221/421/2023-PSE

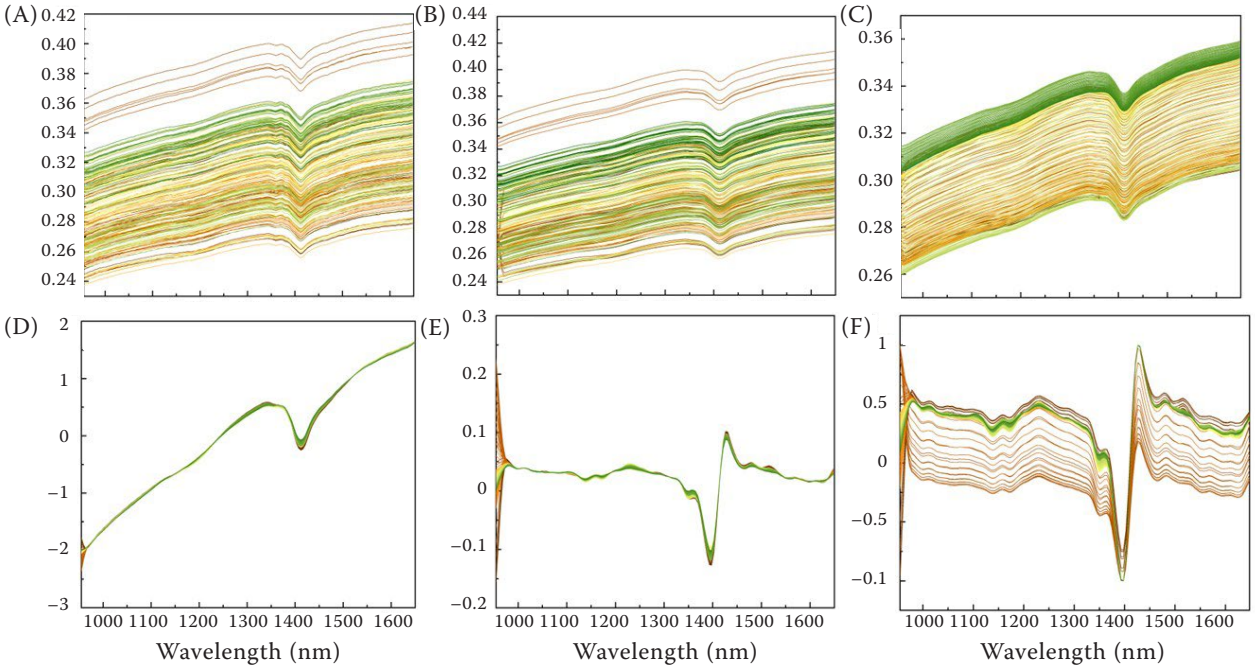

Figure 2. Preprocessed of spectral data. (A) Raw spectra for all the samples; (B) S-G preprocessed spectra for all the samples; (C) S-G + DWT preprocessed spectra for all the samples; (D) S-G + DWT + SNV preprocessed spectra for all the samples; (E) S-G + DWT + SNV + derivative preprocessed spectra for all the samples; (F) S-G + DWT + SNV + derivative + Mapminmax spectra for all the samples

In addition to reducing the dimensions of variables, the model selection substantially impacts its predictive accuracy. PLSR, as a classical linear regression model, has been widely used to estimate the content of the AHN. Because the relationship between AHN and spectral data is rarely linear in nature, linear models may encounter difficulties in modelling. Some studies have shown that nonlinear models are more suitable for dealing with complex nonlinear relationships between AHN and spectral data. For example, Xu et al. (2021) predicted the soil properties for intact soil cores in paddy fields. The result showed that BPNN and LSSVM provide better performance than PLS. In this work, the CARS algorithm was combined with PLSR, BPNN, and LSSVM to construct CARS-PLS, CARS-BPNN, and CARS-LSSVM prediction models, respectively (Liu et al. 2017, 2019, Yang et al. 2020). The three models were run 100 times each. The predictive ability of the models was evaluated using the coefficient of determination of the correction (R_c^2) and the prediction (R_p^2), the root mean square error of the correction set (RMSEC), the root mean square error of the prediction set (RMSEP), and the relative analytical error (RPD). The values of R_c^2 and R_p^2 are between 0 and 1. For RPD, 1.5 < RPD < 2.0 indicates that the model can only make a rough estimate of the high or low content of the samples. 2.0 ≤ RPD < 2.5 indicates that the model has a good predictive ability. And 2.5 ≤ RPD < 3.0 indicates that the model has very good predictive ability (Chen et al. 2022, Li et al. 2022a, 2023). Larger R^2 and RPD and smaller RMSE indicate that the model predicts better.

RESULTS

Statistical analysis of the pattern of change of the AHN. The AHN of the basic soil fertility measured in the laboratory was 47.94 mg/kg. The contents of the AHN at the seedling, jointing, tasseling, filling and mature stages were 31.65–100.84, 28.79–74.05, 44.33–95.61, 54.39–174.42 and 24.42–92.01 mg/kg, respectively. The content ranged from 24.42 to 174.42 mg/kg during the whole stage. As the sampling sites were farmland areas, variations between AHN's maximum and minimum values were substantial due to varying fertiliser application methods and cropping systems in different plots. The average values were 51, 59.06, 73.12, 81.67 and 41.76 mg/kg at the seedling, jointing, tasseling, filling and mature stages,

https://doi.org/10.17221/421/2023-PSE

respectively. In the whole period, the average value was 61.32 mg/kg. The average value of AHN increased from the seedling period to the filling stage. The AHN refers to inorganic nitrogen, which consists of nitrate nitrogen, ammonium nitrogen and soluble organic nitrogen. The base fertiliser used organic fertiliser. After conversion, the organic nitrogen in the organic fertiliser was converted into inorganic nitrogen. At the mature stage, the average value of AHN decreased. As a large amount of nutrients were transferred to the plant for fruit growth at maturity, there was less AHN in the soil.

Sample partition and data preprocessing. This work aimed to construct a predictive model of the AHN in the soil. When developing a model, it is necessary to have calibration sets to create the model and conduct cross-validation. Prediction sets are then utilised to assess the predictive accuracy of the model. The sample set partitioning based on Joint X-Y Distance (SPXY) algorithm can optimise the search for samples and determine the most efficient way to partition them, with the objective of minimising the differences between the resulting subsets (Li et al. 2022a). The SPXY algorithm was used to group all samples (N = 225) into two datasets, of which 169 (75% of the total) and 56 (25% of the total) samples formed the correction and prediction datasets, respectively. The data collected from the spectrometer is noise-prone and requires preprocessing. We used S-G, DWT, SNV, Derivative and Mapminmax algorithms in this work. These algorithms were used to preprocess the collected data. We evaluated the efficiency and accuracy of each algorithm and compared the results. The results indicated that utilising all five algorithms was the most effective method for preprocessing the spectral data. This approach yielded the most precise and efficient outcomes compared to the other preprocessing techniques. The characteristic bands were extracted using the CARS algorithm. Table 2 displays the statistical reference values for the AHN in the soil using the best pretreatment strategy. The similarity between the sample mean of the calibration and prediction indicates the rationality of the dataset partitioning.

Results of feature extraction by CARS. The MCS number was set to 100 times during the optimisation process. Figure 3 displays the optimisation process and the selected optimal wavelengths. This figure illustrates the variation in the number of sample variables, RMSECV, and regression coefficient paths in the subset with the increased number of MCS runs. As shown in the figure, increasing the number of MCS from 1 to 100 resulted in a decrease in the number of collected variables from 148 to 13. The RMSECV exhibited a trend of fluctuating from high to low and back to high again, which decreased to a minimum of 16.7205 mg/kg at 26 sampling intervals. This indicated that information variables unrelated to the AHN were eliminated during the variable selection operation from 1 to 25. The RMSECV began to increase after 26 iterations. It was possible that important variables associated with AHN may have been removed. This could result in an increase in the RMSECV value and a degradation of model efficiency. Regression coefficient path change plots were marked using the vertical lines with asterisks. The optimal subset corresponded to the lowest RMSECV. Finally, the CARS model yielded 13 characteristic wavelengths: 978, 1 025, 1 073, 1 143, 1 223, 1 228, 1 233, 1 337, 1 417, 1 497, 1 501, 1 506, and 1 643 nm. These results demonstrated that the CARS algorithm effectively minimised the number of band inputs and eliminated data redundancy.

Results of modelling. In this work, we utilised the CARS algorithm to extract feature bands, which were modelled separately by combining three algorithms: PLS, BPNN and LSSVM. The entire experiment utilised 225 soil samples, with 169 for modelling and 56 for validating the model. The selected characteristic bands, chosen by the CARS algorithm, were utilised with the PLSR, BPNN, and LSSVM model construction methods to accurately predict the AHN content. The final results of the complete experiment are shown in Table 3. According to the Table 3:

(1) The characteristic bands screened according to the CARS algorithm were entered as input variables into the constructed prediction model of the AHN. The modelling verification accuracy R_p^2 of the three models was ranked in descending order as LSSVM > PLS > BPNN. The LSSVM model exhibited the greatest predictive accuracy among the three established models. The validation accuracies of R_p^2, RMSEP and RPD were 0.8295, 2.95 and 2.42, respectively. Both PLS and BPNN have lower predicting accuracy compared to the LSSVM. The prediction accuracy of the PLS model is in the middle position among the three models; the validation accuracies of R_p^2, RMSEP and RPD were 0.8165, 9.31 and 2.33, respectively. The prediction model built by BPNN had the lowest accuracy. Its model validation accuracies of R_p^2, RMSEP and RPD were 0.7626, 10.43 and 2.05, respectively.

https://doi.org/10.17221/421/2023-PSE

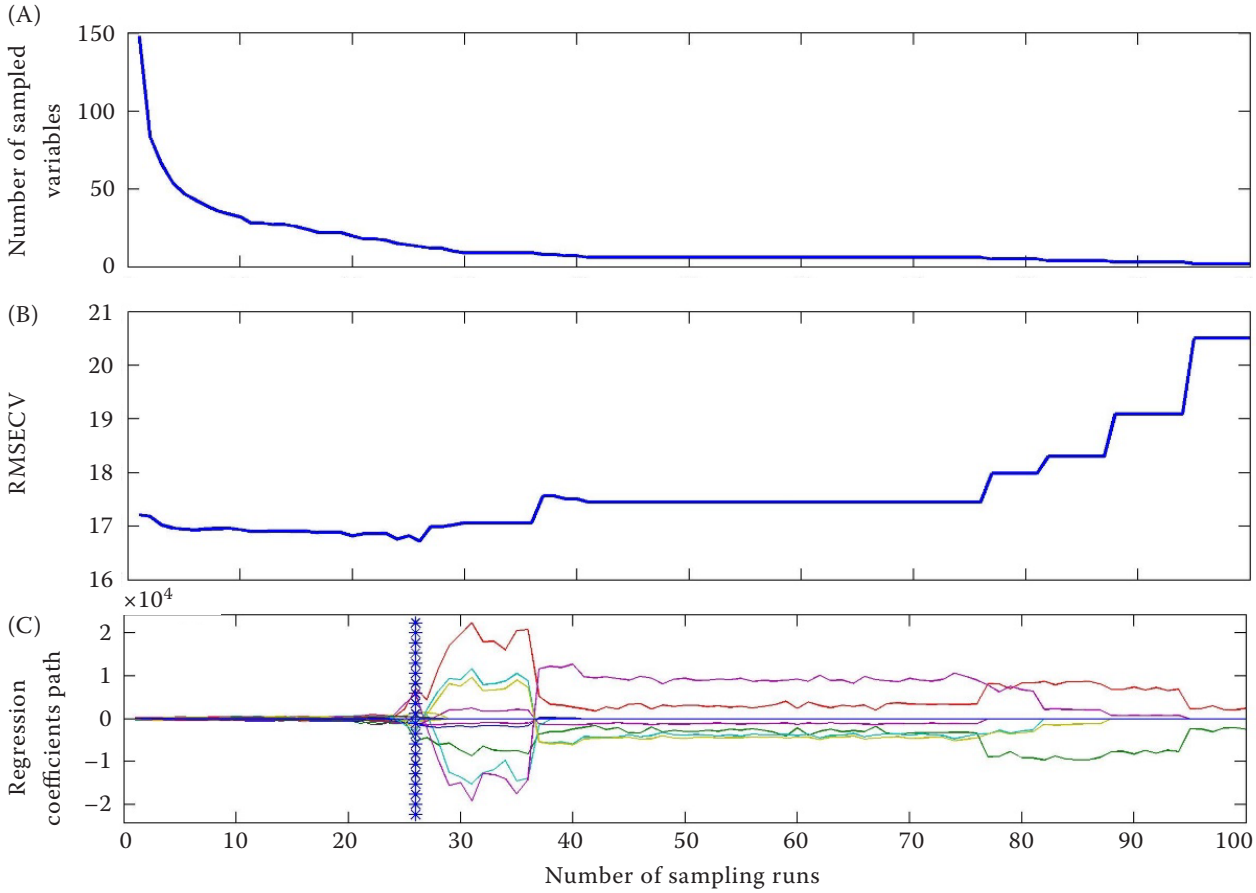

Figure 3. Result of CARS running. Where: (A) refers to the change in the number of sample variables; (B) refers to the variation of root mean squared error of cross-validation (RMSECV), and (C) refers to the change of the regression coefficient path

(2) Comparing the results of these three models, it is clear that the modelling accuracy of the LSSVM is higher than that of the PLS and BPNN models. The RPD of the validation model utilising LSSVM is 2.42. The results indicate that the LSSVM model can approximate AHN's content with great accuracy. Furthermore, the model exhibits an exceptional estimation ability. The LSSVM was the best prediction model.

The CARS-LSSVM model was created with the intention of directly expressing the content of the AHN through prediction performance. The predicted

Table 3. Prediction model of alkali-hydrolysed nitrogen (AHN)

	Model	R_c^2	RMSEC	R_p^2	RMSEP	RPD
FULL	PLSR	0.6584	15.99	0.7911	9.86	2.19
	BPNN	0.4142	19.34	0.6345	12.46	1.65
	LSSVM	0.7681	3.08	0.7974	2.67	2.22
CARS	PLSR	0.6585	15.99	0.8165	9.31	2.33
	BPNN	0.4760	18.69	0.7626	10.43	2.05
	LSSVM	0.7736	3.43	0.8295	2.95	2.42

FULL – full bands; CARS – bands after feature extraction by CARS algorithm; PLSR – partial least squares regression; BPNN – back propagation neural network; LSSVM – least squares support vector machine; R_c^2 – determination coefficient of the correction set; RMSEC – root mean square error of the correction set; R_p^2 – determination coefficient of the prediction set; RMSEP – root mean square error of the prediction set; RPD – ratio performance deviation

https://doi.org/10.17221/421/2023-PSE

and true values of the calibration set (169 samples) and prediction set (56 samples) are shown in Figure 4.

DISCUSSION

In the past, scholars often relied on correlation analysis to examine the correlation between AHN and soil spectral reflectance, including its various mathematical transformations. The high correlation coefficient band was typically considered the sensitive band. Then, more and more scholars adopted the CARS variable optimisation method to filter out the invalid or redundant variables from the full band and select the sensitive bands. In this work, 13 sensitive bands were selected based on the CARS algorithm, representing 8.8% of the total number of bands. This greatly reduces the band information and solves the problem of the large number of bands and heavy computation in the prediction of the AHN. Therefore, the characteristic bands screened by the CARS algorithm coincide with the reflectance spectral response bands of the AHN that have been studied. There are also differences in the optimal subsets of different soil types selected through the CARS algorithm, which selects characterising variables with instability. Vibrational bands of nitrite are detectable in the soil at the wavelength of 978 nm. At 1 025 nm, there is nitrite nitrogen in the soil. In the near-infrared range between 1 060 and 1 080 nm, certain organic compounds which contain carbon-nitrogen bonds display vibrational absorption peaks. There are vibrations and transitions of nitrogen oxides within the 1 070–1 080 nm range that manifest as absorption peaks. The stretching vibration of phosphate (PO_4^{3-}) is observed at 1 143 nm. In the 1 140–1 150 nm range, phosphates exhibit absorption peaks along with some vibrational bands, especially phosphate ion (PO_4^{3-}) or dihydrogen phosphate ion (HPO_4^{2-}). The C-N stretching vibration and C-N-H bending vibration of organic nitrogen compounds in soil are present at wavelengths 1 223, 1 228, and 1 233 nm. The stretching vibration of nitrate was observed at 1 337 nm. In the wavelength range between 1 330 and 1 340 nm, there are some vibrational bands of absorption peaks for nitrate, particularly nitrate ion (NO_3^-) or nitrate dihydrogen ion (HNO_3^-). The C-H and N-H bending vibrations in soil organic nitrogen compounds occur at 1 417 nm. The inorganic nitrogen compounds in the soil exhibit noteworthy absorption properties at wavelengths of 1 497, 1 502, and 1 506 nm. Resonance absorption occurs between the light in the specific wavelength range and molecules of inorganic nitrogen compounds in the soil. This phenomenon arises from the vibration patterns occurring within the molecules. The 1 495–1 510 nm bands fall within the wavelength range of this vibrational absorption.

In this work, the soil samples of different fertility periods in the experimental field of millets were taken as the research object – hyperspectral imaging techniques combined with machine learning algorithms for predictive modelling of the content of the AHN. The results of the work showed that the best model prediction results were obtained by using CARS-LSSVM. The R_c^2, R_p^2, RMSEC, RMSEP and RPD are 0.7736, 3.43, 0.8295, 2.95 and 2.42, respectively. This indicates that the model constructed by LSSVM can be used to predict the content of the AHN. The

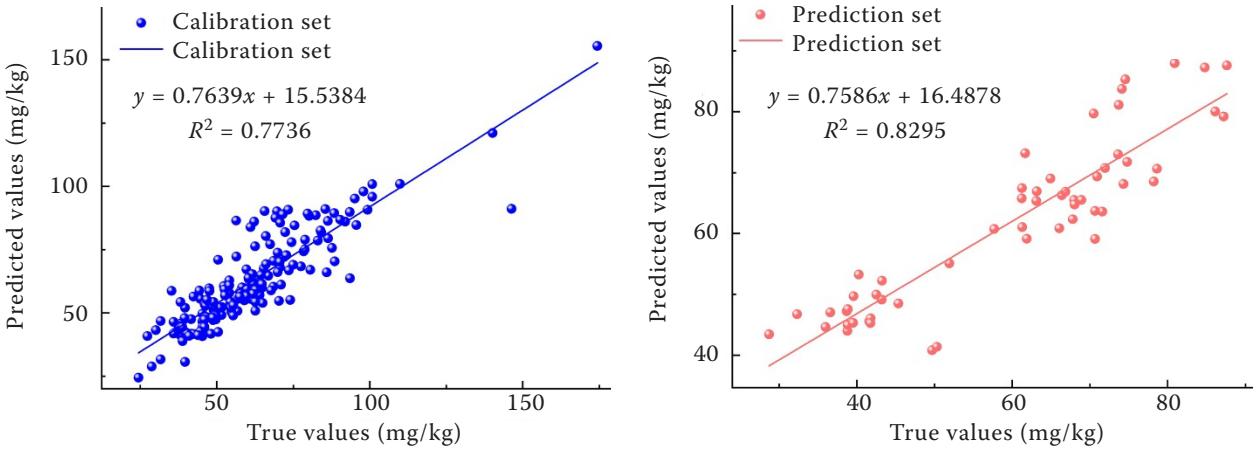

Figure 4. Results of the calibration set and prediction set for the CARS-LSSVM (competitive adaptive reweighted sampling-least squares support vector machine) model of the content of alkali-hydrolysed nitrogen

https://doi.org/10.17221/421/2023-PSE

model has a good predictive ability. The LSSVM is the best estimator among the three machine learning algorithms. This result can provide data support for applying hyperspectral imaging technology to detect the content of the AHN in soil and provide a scientific basis for crop fertiliser management. Meanwhile, the prediction of the content of the AHN can also provide a theoretical reference for the rapid detection of other soil nutrients using hyperspectral imaging technology. The future directions mainly include improving experimental protocols and procedures, using new instruments and techniques to achieve more comprehensive analysis, and learning more accurate and efficient data interpretation and processing methods. Continuous improvements will enhance the analytical capabilities of models. The combination of spectral technology and soil nutrient research can improve the accuracy and repeatability of data and provide a more comprehensive, efficient and accurate soil nitrogen assessment method. This will help to improve the accuracy and efficiency of soil nitrogen assessment.

Acknowledgement. The content of the alkali-hydrolysed nitrogen in the soil was determined in the chemistry laboratory of the Department of Basic Sciences, Shanxi Agricultural University. We thank Professor Huiling Du's team for providing the experimental instruments and reagents.

REFERENCES

Bhadra S., Sagan V., Maimaitijiang M., Maimaitiyiming M., Newcomb M., Shakoor N., Mockler T.C. (2020): Quantifying leaf chlorophyll concentration of sorghum from hyperspectral data using derivative calculus and machine learning. Remote Sensing, 12: 2082.

Chen Z., Ren S., Qin R., Nie P. (2022): Rapid detection of different types of soil nitrogen using near-infrared hyperspectral imaging. Molecules, 27: 2017.

Li C., Chen H., Zhang Y., Hong S., Ai W., Mo L. (2022a): Improvement of NIR prediction ability by dual model optimization in fusion of NSIA and SA methods. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 276: 121247.

Li H., Meng Z., Dang X., Yang P. (2022b): Checkerboard barriers attenuate soil particle loss and promote nutrient contents of soil. Sustainability, 14: 10492.

Li Y., Yang X. (2023): Quantitative analysis of near infrared spectroscopic data based on dual-band transformation and competitive adaptive reweighted sampling. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 285: 121924.

Liu C., Hao G., Su M., Chen Y., Zheng L. (2017): Potential of multispectral imaging combined with chemometric methods for rapid detection of sucrose adulteration in tomato paste. Journal of Food Engineering, 215: 78–83.

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

Liu S., Xiao Z. (2021): Research on hyperspectral inversion of soil alkaline hydrolysis nitrogen content and pH value based on DWD. Journal of Physics: Conference Series, 2079, No. 1, IOP Publishing, 012021.

Liu W., Zhao P., Wu C., Liu C., Yang J., Zheng L. (2019): Rapid determination of aflatoxin B1 concentration in soybean oil using terahertz spectroscopy with chemometric methods. Food Chemistry, 293: 213–219.

Niu J., Chao J., Xiao Y., Chen W., Zhang C., Liu X., Rang Z., Yin H., Dai L. (2017): Insight into the effects of different cropping systems on soil bacterial community and tobacco bacterial wilt rate. Journal of Basic Microbiology, 57: 3–11.

Niu Z., Shi L., Qiao H., Xu X., Wang W., Ma X., Zhang J. (2023): Construction of a hyperspectral estimation model for total nitrogen content in Shajiang black soil. Journal of Plant Nutrition and Soil Science, 186: 196–208.

Pechanec V., Mráz A., Rozkošný L., Vyvlečka P. (2021): Usage of airborne hyperspectral imaging data for identifying spatial variability of soil nitrogen content. ISPRS International Journal of Geo-Information, 10: 355.

Qi H., Paz-Kagan T., Karnieli A., Jin X., Li S. (2018): Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. Soil and Tillage Research, 175: 267–275.

Sun J., Lu X., Mao H., Jin X., Wu X. (2017): A method for rapid identification of rice origin by hyperspectral imaging technology. Journal of Food Process Engineering, 40: e12297.

Tahmasbian I., Xu Z., Boyd S., Zhou J., Esmaeilani R., Che R., Bai S.H. (2018): Laboratory-based hyperspectral image analysis for predicting soil carbon, nitrogen and their isotopic compositions. Geoderma, 330: 254–263.

Tang R., Li X., Li C., Jiang K., Hu W., Wu J. (2022): Estimation of total nitrogen content in rubber plantation soil based on hyperspectral and fractional order derivative. Electronics, 11: 1956.

Xie X., Pu L., Zhu M., Xu Y., Wang X. (2019): Linkage between soil salinization indicators and physicochemical properties in a long-term intensive agricultural coastal reclamation area, Eastern China. Journal of Soil Sediment, 19: 3699–3707.

Xing Z., Du C., Shen Y., Ma F., Zhou J. (2021): A method combining FTIR-ATR and Raman spectroscopy to determine soil organic matter: improvement of prediction accuracy using competitive adaptive reweighted sampling (CARS). Computers and Electronics in Agriculture, 191: 106549.

Xu M., Sun J., Zhou X., Tang N., Shen J., Wu X. (2021): Research on nondestructive identification of grape varieties based on EEMD-DWT and hyperspectral image. Journal Food Science, 86: 2011–2023.

https://doi.org/10.17221/421/2023-PSE

Xu S., Wang M., Shi X., Yu Q., Zhang Z. (2021): Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. Science of The Total Environment, 754: 142135.

Yang H., Liu Y., Xiong Z., Liang L. (2020): Rapid determination of holocellulose and lignin in wood by near infrared spectroscopy and kernel extreme learning machine. Analytical Letters, 53: 1140–1154.

Ye R., Chen Y., Guo Y., Duan Q., Li D., Liu C. (2020): NIR hyperspectral imaging technology combined with multivariate methods to identify shrimp freshness. Applied Sciences, 10: 5498.

Zhang G., Hao H., Wang Y., Jiang Y., Shi J., Yu J., Cui X., Li J., Zhou S., Yu B. (2021): Optimized adaptive Savitzky-Golay filtering algorithm based on deep learning network for absorption spectroscopy. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 263: 120187.

Zhang J., Cheng T., Shi L., Wang W., Niu Z., Guo W., Ma X. (2022): Combining spectral and texture features of UAV hyperspectral images for leaf nitrogen content monitoring in winter wheat. International Journal of Applied Earth Observation and Geoinformation, 43: 2335–2356.

Zimmermann B., Kohler A. (2013): Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. Applied Spectroscopy, 67: 892–902.

Received: October 18, 2023
Accepted: November 11, 2023
Published online: December 13, 2023