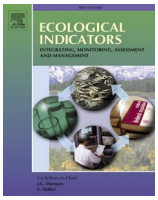
Ecological Indicators 160 (2024) 111843

Contents lists available at ScienceDirect

Ecological Indicators

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



Multi-scale spatial and spectral feature fusion for soil carbon content prediction based on hyperspectral images

Xueying Li [a], Zongmin Li [c], Huimin Qiu [b], Guangyuan Chen [d], Pingping Fan [b,*], Yan Liu [b]

[a] School of Data Science, Qingdao University of Science and Technology, Qingdao 266061, China
[b] Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao 266061, China
[c] College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266590, China
[d] College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

ARTICLE INFO

Keywords:
Soil carbon
Hyperspectral images
Deep learning
Feature extraction

ABSTRACT

Soil carbon content prediction based on hyperspectral images can achieve large-scale spatial measurement, which has the advantages of wide coverage and fast information collection, is more suitable for field data collection. However, the research on soil carbon content prediction based on hyperspectral images mainly focuses on feature extraction of spectral information, ignoring the spatial information, and cannot well reveal the intrinsic structural characteristics of data. Aiming at the lack of spatial features consideration in hyperspectral images, soil carbon content prediction methods based on multi-scale feature fusion are proposed by hyperspectral image. At the same time of extracting spectral features from hyperspectral images, the spatial information is used for the first time and a multi-scale spectral and spatial feature network (SpeSpaMN) is designed. In the SpeSpaMN, the multi-scale spectral feature network (SpeMN) is constructed to extract spectral features, the multi-scale spatial feature network (SpaMN) is constructed to extract spatial features. The two networks are fused by using the complementary relationship between different scale features to achieve soil carbon content prediction based on multi-scale feature fusion. The results showed that SpeSpaMN had the best results compared to other methods, followed by the method of SpeMN. The RPD of Inland, Aoshan Bay and Jiaozhou Bay samples based on SpeSpaMN were increased by 47.36%, 37.96% and 4.30% respectively. This paper can effectively solve the problem of the deep fusion of spatial and spectral features in the soil carbon content prediction by hyperspectral image, so as to improve the accuracy and stability of soil carbon content prediction.

1. Introduction

Hyperspectral imaging technology begins to rise in the field of earth observation in the 1980 s. Imaging spectrometer is a hardware equipment to collect hyperspectral images. With the rapid development of imaging spectrometer in recent decades, hyperspectral imaging technology has been widely used in resource exploration, ecological environment monitoring, land cover classification and target recognition (Hou et al., 2022; Liu et al., 2022; Sun et al., 2022; Yang et al., 2022). In the hyperspectral image (HSI), each pixel contains hundreds of spectral bands. Compared to traditional imaging systems, rich spectral information of HSI helps to better identify surface features and objects (Li et al., 2011).

In the prediction of soil carbon content, visible-near infrared spectroscopy has wide application prospect due to its advantages of non-destructive, rapid, and good reproducibility. The carbon content model based on spectral reflectance measured in the laboratory is with high accuracy (Brown, 2007; Bao et al., 2020). The prediction of soil carbon content by visible-near infrared spectroscopy is a point measurement. Remote sensing technology can achieve large-scale spatial measurement requires remote sensing technology, HSI technology is a typical representative. HSI can obtain a large amount of spatial and spectral information simultaneously (Imani et al., 2020) and has the advantages of wide coverage and fast information acquisition, which is more suitable for field data collection. Landsat and airborne hyperspectral images have been used for monitoring soil regional surface carbon concentration in existing studies (Wang et al., 2022; Jaber et al., 2011). Firstly the original HSI data is preprocessed, then the soil feature information is obtained using the feature variable selection methods, and finally the prediction model of the spectral information in the HSI

* Corresponding author.
E-mail address: fanpp_sdioi@126.com (P. Fan).

https://doi.org/10.1016/j.ecolind.2024.111843
Received 2 June 2023; Received in revised form 16 February 2024; Accepted 1 March 2024
Available online 4 March 2024

and the soil carbon content is established (Wang et al., 2021b; Wang et al., 2021c; Castaldi et al., 2018; Meng et al., 2021; Ribeiro et al., 2021).

The current study of soil carbon content prediction based on HSI only used spectral information. The data source of the soil carbon content model was obtained by acquiring the average spectrum of the selected pixels in the HSI (Dristi et al., 2022; Petri et al., 2023; Michael et al., 2023), which ignored the spatial information of HSI. There was few research on combining spatial and spectral information for soil carbon content prediction. Petri et al. proposed to divide each hyperspectral image into 36 patches in order to provide enough samples for the CNN regressor, which did not make full use of spatial information of HSI (Petri et al., 2023). Multiple underlying spatial feature information (including color, texture, shape features, etc.) combined with spectral information are used for HSI classification (Weng et al., 2020a; Weng et al., 2020b; Xu et al., 2020b; Li et al., 2015; Wei et al., 2021). The fusion of these studies is to extract artificial designed spatial features and spectral features for stack fusion. In order to make full use of hyperspectral feature information, deep learning method is used to extract the spatial and spectral features of HSI. Convolutional neural network (CNN) learns local and abstract features from original spectral data through sparse local connections and weight sharing. The spectral data are continuous, and their order is crucial for predicting soil properties. The same feature peaks appear at different locations in the spectral data may represent different information (Zhang et al., 2020). However, the unsensitivity of CNN to feature positions leads to a decline in model performance (Lecun et al., 1998). Recursive neural network (RNN) is mainly used for sequential problems. It stores historical information over time through feedback connection, and learns the long-term dependence information of sequence data. The spectrum can be regarded as an orderly data.

Yang et al. proposed a jointed CNN and RNN architecture called CCNVR (Yang et al., 2020). It combined the ability of the CNN to extract local and abstract features from the original spectrum and the advantages of various dependencies of the RNN in learning sequence features. Due to the problem of gradient disappearance in RNN, a long-term and short-term memory network (LSTM) was proposed. LSTM could selectively memorize the information provided, which enabled to retain long-term dependent information. Thus the gradient disappearance problem in the RNN was solved, the sequences could be predicted efficiently (Singh et al., 2019). Mou et al. regarded hyperspectral data as sequential data and input to the LSTM, the results demonstrated that spectral data could be used as sequential data (Mou et al., 2017). On this basis, Emile et al regarded the spectral data as sequential data and used LSTM for HSI classification, which also produced good results (Emile et al., 2018). Liu et al. proposed bidirectional-convolutional LSTM to extract the spectral and spatial features for HSI classification (Liu et al., 2017). LSTM could be used to extract spectral information of HSI. In order to better extract spatial sequence information of HSI while retaining the shape of input data, Shi et al. developed convolutional LSTM (ConvLSTM) on the basis of LSTM (Shi et al., 2015). In the ConvLSTM, the full connected layer was substituted with the convolutional layer in the gate to extract the spatial and spectral information of HSI. In order to extract more complete spatial and spectral information from HSI of three-dimensional data, Hu et al. proposed two deep models to capture the feature map of the different channels in the CNN (Hu et al., 2020). Two deep models were the spatial-spectral ConvLSTM 2-D neural network and the spatial-spectral ConvLSTM 3-D neural network. To reduce the trainable parameters and memory requirements of ConvLSTM, Hu et al. proposed a lightweight tensor attention-driven ConvLSTM neural network (Hu et al., 2021).

The research on soil carbon content prediction based on HSI mainly focused on feature extraction of hyperspectral spectral information, ignoring the spatial information of HSI, which failed to reveal the intrinsic structural characteristics of HSI. Existing HSI classification studies described the use of artificial design method to extract spatial features and spectral features for stacking, which could not fully fuse the two features. The feature extraction method based on deep learning could extract spatial and spectral features, especially LSTM, ConvLSTM and other methods could treat hyperspectral data as sequential data. While mining the correlation between spectral bands, the spatial information was extracted to fully combine the spatial and spectral information, which was of great significance for the reference of soil carbon content prediction based on HSI. The current study of soil carbon content prediction based on HSI only uses spectral information, for the problem of how to effectively combine spatial and spectral features in soil carbon content prediction based on HSI, a multi-scale feature fusion method was proposed to achieve soil carbon content prediction. While extracting spectral features from HSI, spatial information of HSI was first used. The complementary relationships between features at different scales were used for fusion to fully mine soil spatial and spectral feature information, which could help to improve the accuracy of soil carbon content prediction in HSI.

2. Experimental and methods

2.1. Experimental samples

The experimental samples was 0–20 cm surface soil collected from Inland, Aoshan Bay and Jiaozhou Bay in Qingdao of China. Fig. 1 was the soil sampling locations map. The Inland sample plot included soil samles of three sites, the land type was mainly sandy loam and silty loam. Inland 1 (60 samples), Inland 2 (44 samples) and Inland 3 (60 samples) were respectively collected from the banks of Licun River, the foothills of Zaoshan Mountain in Qingdao, and the foothills of Fushan Mountain in Qingdao. The Aoshan Bay samples were offshore surrounding soil, the soil type was mainly silty soft clay. Aoshan Bay 1 (64 samples) and Aoshan Bay 2 (56 samples) were respectively collected from the Port East Terminal, and the Wanzimen. The Jiaozhou Bay sample were the soil around the wetland, the soil type was mainly sandy silt, followed by clayey silt. Jiaozhou Bay 1 (78 samples) and Jiaozhou Bay 2 (56 samples) were respectively collected from the Yang River, and the Dagu River. Inland, Aoshan Bay and Jiaozhou Bay represented brown soil, salt soil and tidal soil, and contained mountain soil, wetland soil and coastal soil, which were typical. Brown soil is the most widely distributed and extensive soil type in the inland of Qingdao, mainly distributed in mountainous hills and piedmont plains. Therefore, collecting mountain soil and riverside soil can represent the inland soil of Qingdao.

2.2. Hyperspectral image acquisition and chemical analysis

Dualix Spectral Imaging portable hyperspectral camera GaiaField Pro-V10 was used to obtain HSI of soil samples. The spectral range was 400–1000 nm, the spectral sampling interval was 3.2 nm, the number of spectral channels at Inland samples was 176, and the number of spectral channels at Aoshan Bay and Jiaozhou Bay samples was 360. The soil samples were placed into a circular box of 4 cm in diameter and were arranged on the black plate in order, as shown in Fig. 2. The hyperspectral camera was placed on a tripod with a vertical distance of 1 m from the soil samples, and the soil hyperspectral collection system was established for photographing. The HSI of soil samples collected by the hyperspectral camera were cut by ENVI 5.3 with a rectangular figure of 30 * 30 pixels to obtain the HSI of the soil samples from three sample plots. For the convenience of seeing the soil spectrum of each sample plot, the hyperspectral data of each soil sample was averaged. The average reflectance spectrum of soil samples from inland, Aoshan Bay, and Jiaozhou Bay were shown in Fig. 2.

Soil carbon content was determined by elemental analyzer Perkin-Elmer 2400. The statistics of measured soil carbon content in three sample plots were shown in Table 1.

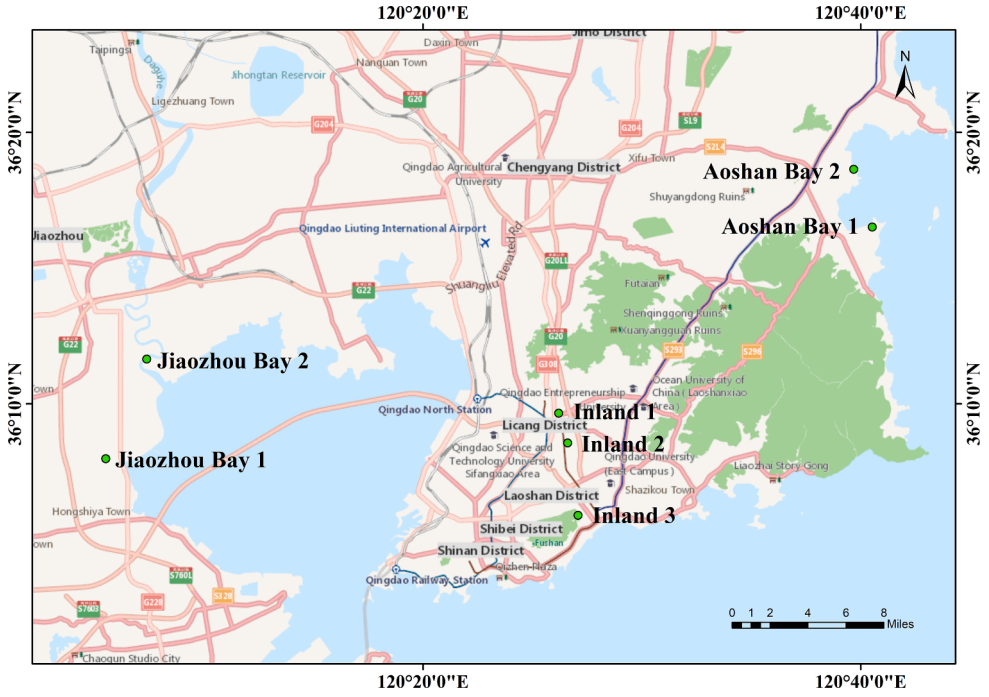


Fig. 1. Soil sampling locations map.

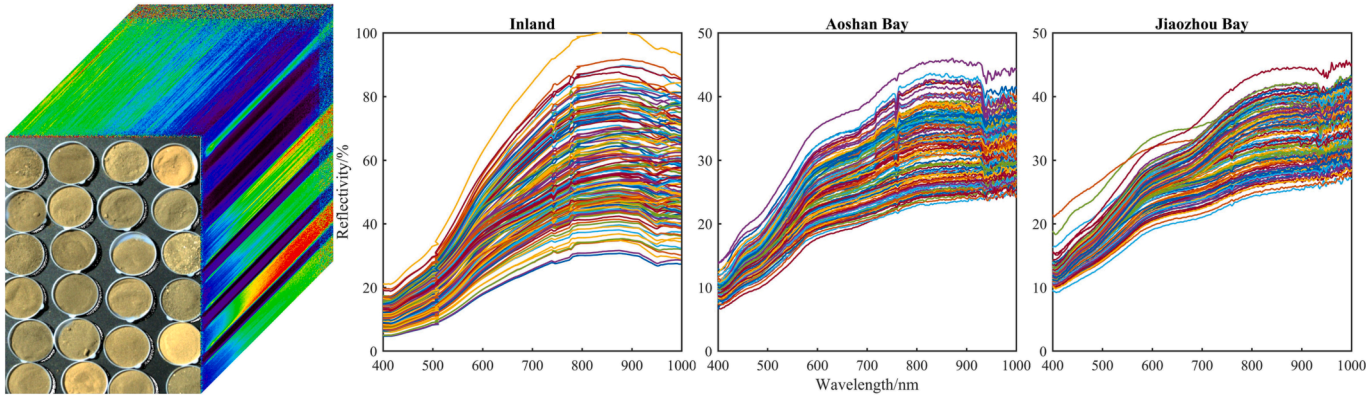


Fig. 2. Example of collecting hyperspectral images of soil samples and the average reflectance spectrum of soil samples from inland, Aoshan Bay, and Jiaozhou Bay.

Table 1
The statistics of measured soil carbon content in three sample plots.

Sample plot	Max/(g·kg^{-1})	Min/(g·kg^{-1})	Mean/(g·kg^{-1})	SD/(g·kg^{-1})
Inland	13.400	1.904	6.369	3.549
Aoshan Bay	9.407	0.997	4.189	2.506
Jiaozhou Bay	15.805	2.670	6.154	2.111

2.3. Methods

2.3.1. Convolutional long short-term memory

RNN is a network structure that consists of multiple loop units connected into a chain. RNN considers the input information from the previous moment, memorizes the information, and computes the current output. Therefore, RNN is very effective for data with sequence properties. Although RNN can effectively handle the sequence deep learning model of long-term dependencies in sequence data (Faroque et al., 2021), it can easily lead to gradient disappearance and explosion problems when processing long-term data (Wang et al., 2021). This is due to the weight matrix cyclic multiplication of RNN, and same functions combined multiple times leads to extreme nonlinear behavior. To solve this problem, the LSTM network is proposed (Hochreiter et al., 1997). LSTM uses the gating mechanism and memory units to effectively capture the contextual information of adjacent data by sequence. By replacing periodic hidden nodes with memory cells as accumulators of state information, the control gates retrieve, update, and clean data from the cells. They allow for the regulation of the information flow. The gradients can be moved through multiple time steps without exploding or disappearing gradients (Faroque et al., 2021).

ConvLSTM is a variant of LSTM. Unlike LSTM, ConvLSTM combines convolution and LSTM to extract the context information of multidimensional data, which overcomes the problem of restricting the shape of the input data in LSTM, and achieves the input-to-state and state-to-state transition through convolution (Hu et al., 2020). Compared to the CNN, ConvLSTM maintains the gate control mechanism. It can not only realize the data transmission and processing between layers, but also perform the data transmission and processing within the layer. The structure of ConvLSTM network is shown in Fig. 3. The calculation formula of ConvLSTM is expressed as:

$$i_t = \sigma(W_{xi} * \mathscr{X}_t + W_{hi} * \mathscr{H}_{t-1} + W_{ci} \otimes \mathscr{C}_{t-1} + b_i) \quad (1)$$

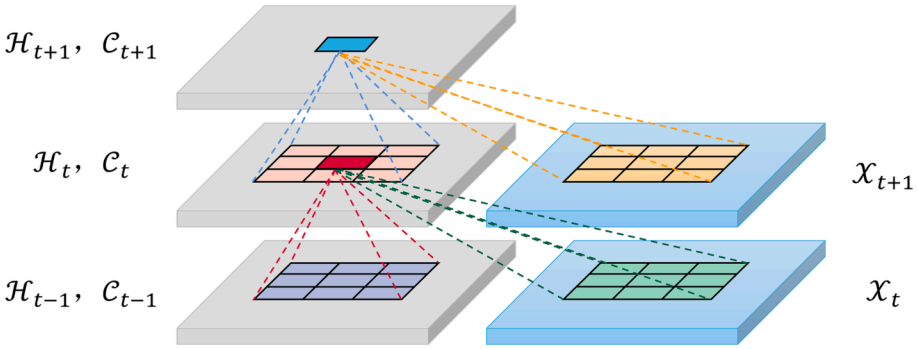


Fig. 3. Structure of ConvLSTM network.

$$f_t = \sigma\left(W_{xf} * \mathscr{X}_t + W_{hf} * \mathscr{H}_{t-1} + W_{cf} \otimes \mathscr{C}_{t-1} + b_f\right) \quad (2)$$

$$c_t = f_t \otimes \mathscr{C}_{t-1} + i_t \otimes tanh(W_{xc} * \mathscr{X}_t + W_{hc} * \mathscr{H}_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo} * \mathscr{X}_t + W_{ho} * \mathscr{H}_{t-1} + W_{co} \otimes \mathscr{C}_t + b_o) \quad (4)$$

$$h_t = o_t \otimes tanh(c_t) \quad (5)$$

Where, $\mathscr{X}_t$ is the multidimensional input of the current cell. $\mathscr{H}_{t-1}$ and $\mathscr{C}_{t-1}$ is the output and status of the previous unit in ConvLSTM. i_t, f_t, c_t and h_t represents the output of input gate, forget gate, output gate, and memory cell at time t in ConvLSTM. $W_{\bullet i}$ and b_i are the weight and deviation of input gate. $W_{\bullet f}$ and b_f are the weight and deviation of forget gate. $W_{\bullet c}$ and b_c are the weight and deviation of memory cell. $W_{\bullet o}$ and b_0 are the weight and deviation of output gate. $tanh(\bullet)$ and $\sigma(\bullet)$ is activation function. $\otimes$ is Hadama product. * is convolution operation.

2.3.2. *Multi-scale spatial and spectral feature fusion algorithm*

In HSI, low-level features have higher resolution and contain more detailed information, but they have lower semantics and more noise. High-level features have stronger semantic information, but their resolution is low and their ability to perceive details is worse. Extracting the hyperspectral features at different scales will present different feature information. Multi-scale fusion can utilize the complementary relationship among features at different scales to fuse, and multiple features are fused to improve the learning ability of the model. How to fully combine and excavate the spatial and spectral information of HSI, fuse the feature information at different scales, and improve the prediction accuracy of soil carbon content, is the key to apply HSI for the quantitative analysis of soil carbon content. There is a correlation between spectral bands, and long-term dependency relationship among spectral bands can be established using LSTM. There are not only spectral information but also spatial information in HSI. Combined with ConvLSTM to extract spatial feature information in HSI to provide effective features for subsequent model building. Multi-scale spectral feature network is constructed based on LSTM and multi-scale spatial feature network is constructed based on ConvLSTM. Then two networks are combined to establish the multi-scale spectral and spatial feature network structure to predict soil carbon content.

(1)Multi-scale spectral feature network.

Spectral feature extraction of HSI plays an important role in subsequent model establishing. Feature extraction of the original data at different scales will present different feature information. Multi-scale feature fusion can improve the complementarity of features. The method of implementing multi-scale hyperspectral spectral information is to diversify the scale of LSTM input data. The spectral data of each sample contains multiple wavelength points, achieving multi-scale effects from the perspective of wavelength division. In order to obtain multiple spectral bands located in different wavelength ranges, the grouping strategy needs to be repeated multiple times (Wang et al., 2021a). The simplest method is to continuously halve the spectral vector and input the results into different LSTMs (Wang et al., 2021a). By merging the outputs of all LSTMs, the final spectral features are obtained, and achieved multi-scale spectral data segmentation. Multi-scale spectral feature includes both global and local information. The specific calculation formula for the spectral data division of each scale is as follows:

$$X_{raw} = [x_1, x_2, \cdots, x_n] \quad (6)$$

$$A(j) = \left[x_{\frac{p}{m}*(j-1)+1}, x_{\frac{p}{m}*(j-1)+2}, \cdots, x_{\frac{p}{m}*j}\right] \quad (7)$$

Where, X_{raw} is original spectral data, a total of p wavelength points. $A(j)$ is the spectral data contained in the j th segmented region after being divided into m parts, As shown in Fig. 4.

Multi-scale spectral feature network (SpeMN) takes the spectral vectors of multiple scales as input to obtain the spectral features of multiple scales through the LSTM network, and then adds the features obtained through the full connection (FC) layer. The added features of all the scales are through the full connection layer, and then are regressed to obtain the final prediction value. Fig. 5 is the structure of multi-scale spectral feature network.

(2)Multi-scale spatial feature network.

HSI contains not only spectral information, but also spatial information. Efficient extraction of spatial information features can further improve the accuracy of the subsequent models. In the HSI, low-level features have higher resolution and contain more detailed information, but they have lower semantics and more noise. High-level features have stronger semantic information, but their resolution is low and their ability to perceive details is poor. Therefore, multi-scale spatial information is used to realize the complementarity among features at different scales, a multi-scale spatial feature network (SpaMN) is constructed. The principle of designing a network is to ensure hierarchical learning of the main branches while obtaining spatial sequence information, extract multi-scale features by being segmented into multiple parts in the spatial dimension. The network structure is shown in Fig. 6.

In HSI, the dilated convolution operation is performed before the convolution operation is performed. The dilated convolution increases the receptive field while ensuring the resolution of the feature mapping to obtain more feature information. The obtained dilated convolution features are subjected to multiple convolution operations. Each convolution operation is as the feature information of different scales, thus obtaining multi-scale hyperspectral feature information from low to high level. Assuming that the HSI are the data of $H_x \times H_y \times H_b$, where H_x is the horizontal axis length of HSI, H_y is the vertical axis length of HSI, H_b is the wavelength length of HSI. The HSI is divided into multiple

Fig. 4. Spectral data division diagram of each scale.

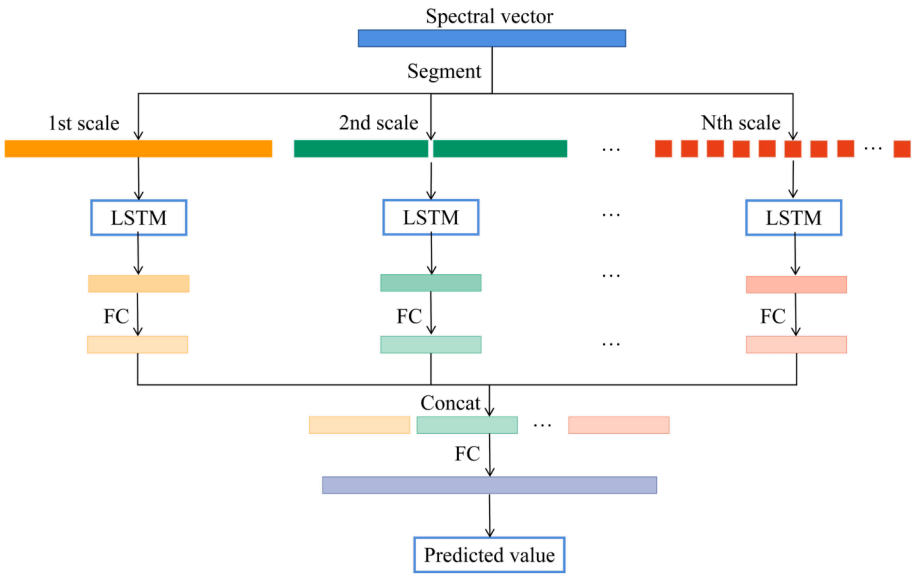


Fig. 5. Structure of multi-scale spectral feature network.

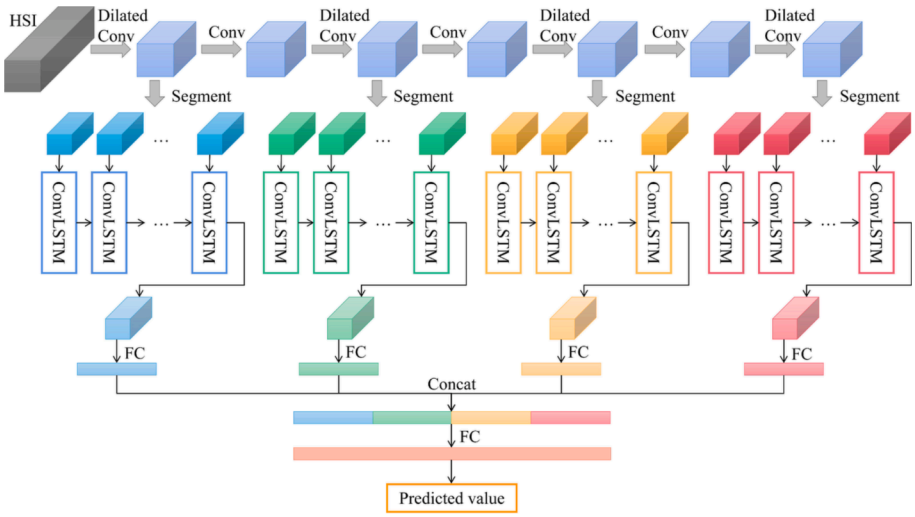


Fig. 6. Structure of multi-scale spatial feature network.

blocks of size $p_x \times p_y \times H_b$, a total of $\frac{H_x}{p_x} \bullet \frac{H_y}{p_y}$ blocks. $\frac{H_x}{p_x} \bullet \frac{H_y}{p_y}$ blocks with size $p_x \times p_y \times H_b$ are input into the ConvLSTM network respectively and the features for each block are obtained. Then each block feature is input into the next ConvLSTM network to obtain the hyperspectral spatial feature information. Take a HSI of size $6 \times 6 \times H_b$ as an example, four blocks with a size of $3 \times 3 \times H_b$ are divided. The operation diagram is shown in Fig. 7. The spatial information at different scales is separately fed into the FC layer to obtain one-dimensional data at different scales, which is added and used for regression prediction.

(3)Multi-scale spectral and spatial feature network.

Multi-scale spectral and spatial feature network (SpeSpaMN) combines SpeMN and SpaMN to obtain hyperspectral spectral and spatial information respectively, and performs regression on spatial and spectral information fusion to obtain predicted values. The structure diagram of SpeSpaMN is shown in Fig. 8. The spectral values of all pixel points in HSI are averaged to obtain the mean spectral vector. The spectral vector is fed into the SpeMN to obtain the spectral feature F_{spec}. HSI is fed directly into the SpaMN to obtain spatial feature information F_{spat}. The spectral feature F_{spec} and spatial features F_{spat} are directly concatenated to obtain the spatial and spectral features F, the calculation formula is as

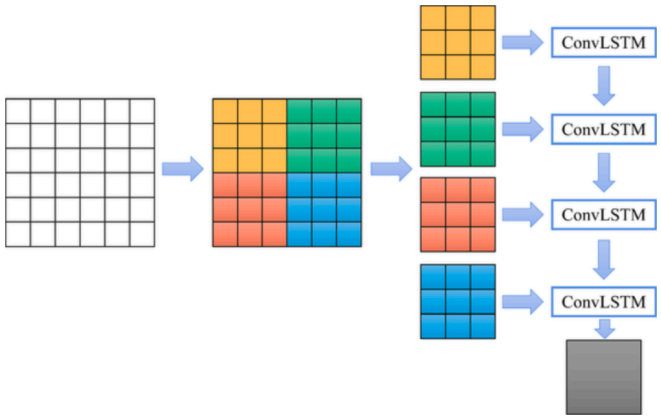


Fig. 7. Hyperspectral spatial feature segmentation diagram.

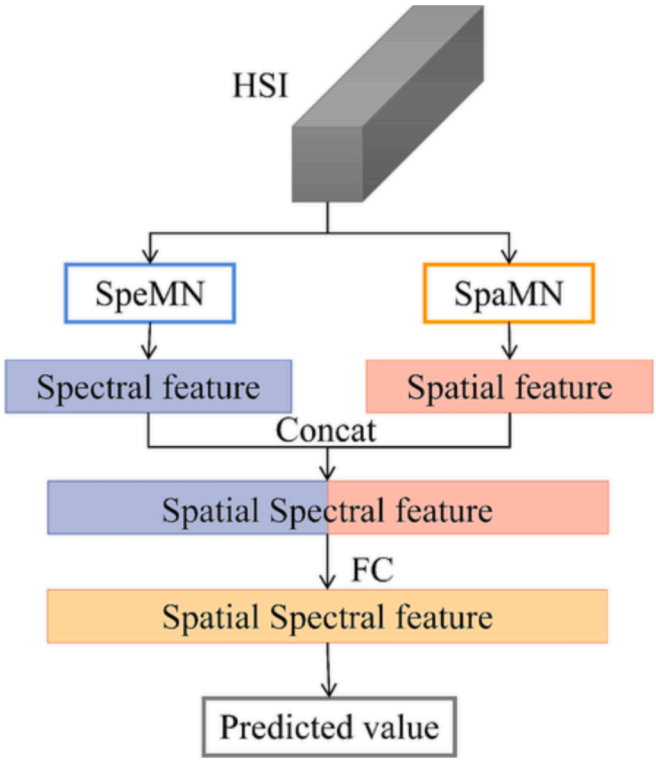


Fig. 8. Structure of multi-scale spectral and spatial feature network.

follows:

$$F = Concat(F_{spec}, F_{spat}) \tag{8}$$

Where, $Concat(.)$ is the concatenation function. The spatial and spectral features F are through the FC layer and then regressed to obtain the predicted value. Both multi-scale spectral feature information and different scales of spatial context features in HSI are extracted by SpeSpaMN. The concatenation method can effectively combines the spatial and spectral features.

Smooth L_1 loss function is used for regression to obtain the predicted value. To fully consider the loss of each network feature, the multiple feature loss is added to calculate the total loss for optimization. $Loss_{spe}$ is the loss of the SpeMN, $Loss_{spa}$ is the loss of the SpaMN. $Loss_{spespa}$ is the loss of the SpeSpaMN. $Loss$ is the total loss by the sum of these three losses, the calculation formula is as follows.

$$Loss = Loss_{spe} + Loss_{spa} + Loss_{spespa} \tag{9}$$

Algorithm: Training SpeSpaMN for soil carbon content prediction based on HSI

Input: The training HSI data X_{train}, measured values of soil carbon content Y_{train}, the test HSI data $X_{predict}$, the number of training epochs E
Output: The predicted values of soil carbon content $Y_{predict}$
1: **Begin**
2: Parameter setting and weights initialization
3: Input the training HSI data X_{train} into the SpeSpaMN
4: **While** step $\leq$ E do
5: Train the model by optimizing the loss function Loss (9)
6: Update the parameters of SpeSpaMN by minimizing the loss
7: **End while**
8: Calculate the carbon content of the test HSI data $X_{predict}$
9: **Return** the predicted values of soil carbon content Y_p
10: **End**

2.3.3. Sample partitioning

The method according to the serial number order, the kennard-stone (KS) method and the sample set partitioning based on joint x-y distance (SPXY) method are used for sample partitioning. The method according to the serial number order selectes the required number of samples in order from the starting point of sample selection for modeling (Li et al., 2023). The KS method selectes samples for calibration set based on the euclidean distance of the spectrum (Wei et al., 2020). The SPXY method is developed on the basis of the KS method, which considered not only spectral information but also carbon content in calculating inter-sample distance (Wei et al., 2020).

2.3.4. Evaluation criteria

The coefficient of determination (R2), root mean square error (RMSE) and relative percent deviation(RPD) are used as the evaluation criteria to measure the quality of the model. The closer the R2 is to 1, the smaller the RMSE, the larger the RPD, and the better the model. R_c^2 and RMSEP are the coefficient of determination and the root mean square error of the calibration set, R_p^2 and RMSEP are the coefficient of determination and the root mean square error of the prediction set.

3. Results and discussion

3.1. Model parameter setting

(1)Parameter setting of multi-scale spectral feature network.

The HSI of each sample was taken the mean value to obtain the spectral vector. The spectral vectors were normalized, and the normalized range was to [-1,1]. The spectral vectors were divided into four scales, including spectral bands/8, spectral bands/4, spectral bands/2 and all spectral bands, and then input into LSTM 1, LSTM 2, LSTM 3, and LSTM 4 networks, respectively. The number of hidden layer neurons in the four LSTM networks was set to 128. The number of FC layers and the output size were both set to 128. The activation function is the nonlinear mapping function ReLU. The parameters of multi-scale spectral feature network was shown in Table 2.

In HSI of Inland, the spectral bands of each pixel were 176, and the input size of LSTM 1, LSTM 2, LSTM 3, and LSTM 4 networks was 22, 44, 88, and 176, respectively. In HSI of Aoshan Bay and Jiaozhou Bay, the spectral bands of each pixel were 360, and the input size of LSTM 1, LSTM 2, LSTM 3, and LSTM 4 networks was 45, 90, 180, and 360, respectively. The learning rate of the three sample plots models was set to 1e-2, the weight decay was set to 1e-6, the adaptive moment estimation (Adam) was adopted for optimization, and the batch-size was set to 16. The number of iterations was set from 10 to 200 with an interval of 10 to calculate R_p^2, RMSEP and RPD of Inland, Aoshan Bay and Jiaozhou Bay, respectively. Prediction evaluation results for different iterations of SpeMN was shown in Fig. 9. In the SpeMN, when the number of iterations in the Inland, Aoshan Bay, and Jiaozhou Bay soil samples were 50, 70, and 20, respectively, the R_p^2 and RPD values were the largest, and the RMSEP was the smallest, with the best prediction effect.

(2)Parameter setting of multi-scale spatial feature network.

The HSI of each sample was normalized, the normalized range was to [-1,1]. The normalized hyperspectral data were input into four dilated

Table 2
The parameters of multi-scale spectral feature network.

Layer	Input size	Hidden layer size	Output size	Activation
LSTM1	Spectral bands/8	128	/	None
LSTM2	Spectral bands/4	128	/	None
LSTM3	Spectral bands/2	128	/	None
LSTM4	Spectral bands	128	/	None
Full connection	128	/	128	ReLU

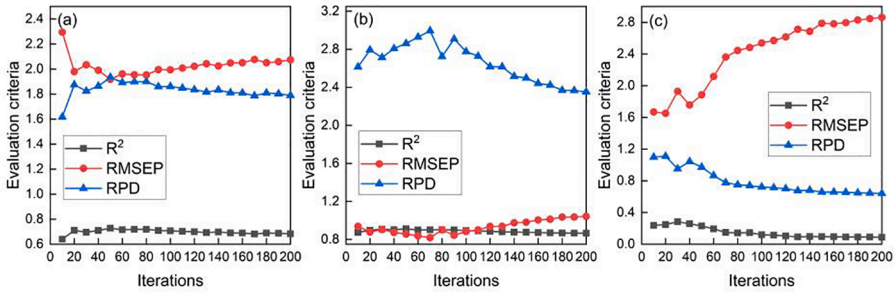


Fig. 9. Prediction evaluation results for different iterations of multi-scale spectral feature network. (a) Inland, (b) Anshan Bay, (c) Jiaozhou Bay.

convolution layers and convolution layers. The number of convolution kernels was set to 32, the convolution kernel size was set to 3 × 3, the stride was set to 1. The padding and dilation rate of convolution layer were set 1. The padding and dilation rate of dilated convolution layer were set 2. The features obtained from the dilated convolution layers and the convolution layers were input the ConvLSTM layers. The parameters of The ConvLSTM layer were consistent with those of convolution layer. The number of FC layers and the output size were both set to 128. The activation function was the nonlinear mapping function ReLU. The parameters of multi-scale spatial feature network was shown in Table 3.

The learning rate of the three sample plots models was set to 1e-2, the weight decay was set to 1e-6, the stochastic gradient descent (SGD) was adopted for optimization, and the batch-size was set to 16. The number of iterations was set from 10 to 200 with an interval of 10 to calculate R_p^2, RMSEP and RPD of three sample plots respectively. Prediction evaluation results for different iterations of SpaMN was shown in Fig. 10. In the SpaMN, when the number of iterations in the Inland, Aoshan Bay, and Jiaozhou Bay soil samples were 160, 100, and 80, respectively, the R_p^2 and RPD values were the largest, and the RMSEP was the smallest, with the best prediction effect.

To select the number of segmented blocks and dropout probability in the three sample plots, the number of segmented blocks was set to 1, 2, 3, 4, and the dropout probability ranged from 0 to 0.9 with intervals of 0.1. Prediction results of different number of segmented blocks and dropout rates in Inland, Aoshan Bay and Jiaozhou Bay were shown in the Table 4, Table 5 and Table 6, respectively. When the number of segmented blocks was 3 and dropout rate was 0.6 in Inland, the prediction effect was best. When the number of segmented blocks was 3 and dropout rate was 0.9 in Aoshan Bay, the prediction effect was best. When the number of segmented blocks was 2 and dropout rate was 0.8 in Jiaozhou Bay, the prediction effect was best.

3.2. Experimental comparison between multi-scale spatial and spectral feature fusion method and traditional methods

The soil samples of Inland, Aoshan Bay and Jiaozhou Bay were divided into calibration set and prediction set in ratio of 2:1 according to the serial number order. After normalizing the hyperspectral data, the spectral features were extracted by SpeMN, the spectral and spatial features were extracted by SpeSpaMN, and the soil carbon content models in the three plots were established. Uninformative variable elimination (UVE), successive projections algorithm (SPA), genetic algorithm (GA), pearson correlation coefficient (PCC), competitive adaptive reweighting sampling (CARS) and random frog (RF) were the commonly used spectral features extracted methods (Bao et al., 2020, Vohland et al., 2017; Xu et al., 2020a). Gray level co-occurrence matrix (GLCM), Gabor and local binary patterns (LBP) were the commonly used low-level spatial features extracted methods (Chu et al., 2021). Therefore, six spectral features extracted method, three low-level spatial features extracted methods combined with full spectrum were used to establish soil carbon content model by PLSR, respectively. The prediction results of multiple feature extraction methods in three sample plots were compared and were shown in the Table 7, Table 8 and Table 9, respectively. Fitting results of soil carbon content measured values and predicted values by multiple feature extraction methods in Inland were shown in Fig. 11, Fig. 12 and Fig. 13.

In the soil carbon content prediction results by multiple feature extraction methods in Inland, based on the analysis of RMSEP and RPD, except for PCC and RF, other feature extraction methods were superior to full-spectrum prediction results. Based on the analysis of R_p^2, the R_p^2 of SPA, PCC, SpeMN, and SpeSpaMN were higher than that of full spectrum. Based on the analysis of the three evaluation criteria, the R_p^2, RMSEP and RPD of SPA, SpeMN and SpeSpaMN were better than the full spectrum. The prediction accuracy of SPA was slightly higher than that of the full spectrum, improving by 0.008 for R_p^2, decreasing 0.201 for RMSEP, and increasing 0.14 for RPD. The prediction accuracy of SpeMN was higher than that of SPA, improving 0.059 for R_p^2, decreasing 0.501 for RMSEP, and increasing 0.401 for RPD compared with the full spectrum. SpeSpaMN had the highest prediction accuracy. The prediction accuracy of these methods from best to worst was following: SpeSpaMN, SpeMN, CARS, SPA, UVE, GLCM, LBP, Gabor, GA, Full-spectrum, RF and PCC. Compared with the full spectral prediction results, R_p^2 was improved by 0.132, RMSEP was decreased by 0.777, and RPD was increased by 0.726. In contrast with existing methods that extracting only spectral features and only spatial features, extracting hyperspectral feature information through SpeMN and SpeSpaMN could improve the accuracy of the soil carbon content model in Inland. In particular, SpeSpaMN obtained a prediction result with an RPD value of 2.259, which was improved from only rough carbon content prediction to accurate carbon content prediction.

In the soil carbon content prediction results by multiple feature extraction methods in Aoshan Bay, based on the analysis of the three evaluation criteria, all the feature extraction methods were better than the full spectral prediction except GA and PCC. Among the extracting spectral features methods, RF had the highest prediction accuracy.

Table 3
The parameters of multi-scale spatial feature network.

Layer	Kernel number	Kernel size	Stride	Padding	Dilation rate	Activation
Convolution	32	3 × 3	1	1	1	ReLU
Dilated Convolution	32	3 × 3	1	2	2	ReLU
ConvLSTM	32	3 × 3	1	1	1	ReLU
Full connection	128	/	/	/	/	None

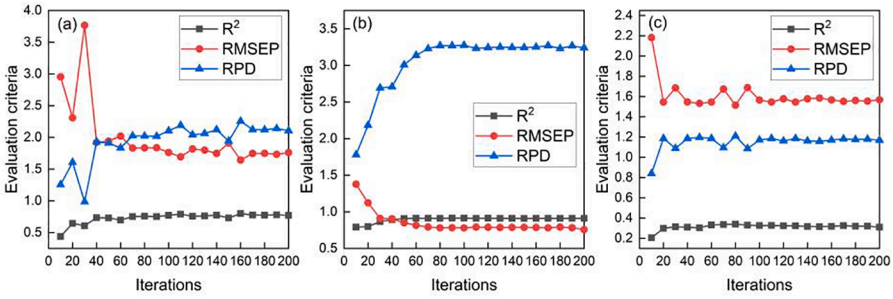


Fig. 10. Prediction evaluation results for different iterations of multi-scale spatial feature network. (a) Inland, (b) Anshan Bay, (c) Jiaozhou Bay.

Table 4
Prediction results of different number of segmented blocks and dropout rates in Inland.

Evaluation criteria	Division number	Dropout probability									
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
R_p^2	1	0.629	0.703	0.713	0.671	0.667	0.660	0.728	0.690	0.707	0.721
	2	0.703	0.702	0.678	0.676	0.701	0.676	0.704	0.725	0.700	0.737
	3	0.779	0.757	0.737	0.764	0.745	0.791	0.801	0.775	0.761	0.732
	4	0.753	0.740	0.735	0.719	0.726	0.734	0.731	0.729	0.723	0.696
RMSEP	1	2.247	2.004	1.977	2.115	2.126	2.154	1.925	2.049	2.002	1.944
	2	2.005	2.009	2.094	2.093	2.011	2.095	1.999	1.927	2.012	1.885
	3	1.730	1.814	1.886	1.789	1.859	1.689	1.644	1.757	1.808	1.904
	4	1.826	1.880	1.898	1.948	1.925	1.898	1.909	1.916	1.934	2.029
RPD	1	1.651	1.851	1.876	1.754	1.745	1.722	1.927	1.811	1.853	1.909
	2	1.851	1.846	1.772	1.772	1.844	1.770	1.856	1.925	1.844	1.968
	3	2.144	2.045	1.967	2.074	1.996	2.196	2.257	2.111	2.052	1.948
	4	2.032	1.974	1.955	1.904	1.927	1.954	1.943	1.936	1.918	1.829

Table 5
Prediction results of different number of segmented blocks and dropout rates in Aoshan Bay.

Evaluation criteria	Division number	Dropout probability									
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
R_p^2	1	0.837	0.838	0.851	0.846	0.799	0.835	0.869	0.861	0.853	0.897
	2	0.878	0.900	0.898	0.885	0.895	0.904	0.908	0.905	0.911	0.910
	3	0.904	0.911	0.908	0.908	0.910	0.911	0.914	0.911	0.913	0.915
	4	0.899	0.903	0.911	0.910	0.910	0.911	0.911	0.910	0.915	0.912
RMSEP	1	1.049	1.053	0.993	0.990	1.187	1.094	0.989	0.970	0.998	0.855
	2	0.909	0.829	0.825	0.930	0.866	0.816	0.803	0.808	0.759	0.757
	3	0.798	0.786	0.785	0.785	0.797	0.781	0.781	0.775	0.748	0.741
	4	0.836	0.815	0.815	0.780	0.796	0.787	0.789	0.786	0.783	0.776
RPD	1	2.337	2.328	2.469	2.478	2.066	2.242	2.480	2.529	2.458	2.869
	2	2.698	2.960	2.971	2.637	2.830	3.007	3.054	3.035	3.232	3.240
	3	3.204	3.256	3.257	3.257	3.208	3.275	3.274	3.299	3.279	3.311
	4	3.060	3.138	3.137	3.145	3.214	3.249	3.241	3.253	3.268	3.295

Table 6
Predictionresults of different number of segmented blocks and dropout rates in Jiaozhou Bay.

Evaluation criteria	Division number	Dropout probability									
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
R_p^2	1	0.232	0.145	0.230	0.233	0.200	0.199	0.188	0.242	0.168	0.325
	2	0.290	0.314	0.291	0.298	0.301	0.311	0.304	0.335	0.345	0.338
	3	0.329	0.332	0.318	0.330	0.324	0.330	0.332	0.339	0.344	0.344
	4	0.333	0.338	0.318	0.327	0.333	0.326	0.341	0.344	0.337	0.336
RMSEP	1	1.752	1.963	1.801	1.675	1.751	1.761	1.777	1.716	1.989	1.605
	2	1.549	1.516	1.547	1.539	1.541	1.538	1.535	1.489	1.480	1.493
	3	1.514	1.509	1.519	1.510	1.540	1.516	1.535	1.519	1.515	1.527
	4	1.624	1.632	1.657	1.658	1.659	1.649	1.633	1.631	1.644	1.643
RPD	1	1.046	0.934	1.018	1.094	1.047	1.041	1.032	1.068	0.922	1.142
	2	1.184	1.210	1.185	1.191	1.190	1.192	1.195	1.231	1.239	1.228
	3	1.211	1.215	1.207	1.214	1.191	1.209	1.194	1.207	1.210	1.201
	4	1.129	1.124	1.107	1.106	1.105	1.112	1.123	1.124	1.116	1.116

Table 7
Prediction results of soil carbon content by multiple feature extraction methods in Inland.

Method	R_c^2	RMSEC	R_p^2	RMSEP	RPD
Full-spectrum	0.635	2.096	0.669	2.419	1.533
UVE	0.648	2.056	0.633	2.245	1.652
SPA	0.687	1.941	0.677	2.218	1.673
GA	0.632	2.104	0.604	2.387	1.554
PCC	0.634	2.096	0.673	2.577	1.439
CARS	0.739	1.772	0.661	2.188	1.696
RF	0.687	1.939	0.553	2.507	1.480
GLCM	0.672	1.987	0.661	2.323	1.597
Gabor	0.660	2.021	0.660	2.378	1.560
LBP	0.701	1.895	0.597	2.339	1.586
SpeMN	0.746	1.584	0.728	1.918	1.934
SpeSpaMN	0.832	1.427	0.801	1.642	2.259

Table 8
Prediction results of soil carbon content by multiple feature extraction methods in Aoshan Bay.

Method	R_c^2	RMSEC	R_p^2	RMSEP	RPD
Full-spectrum	0.869	0.943	0.841	1.022	2.400
UVE	0.824	1.095	0.860	0.932	2.630
SPA	0.749	1.306	0.868	0.948	2.588
GA	0.595	1.659	0.592	1.660	1.477
PCC	0.826	1.086	0.833	1.048	2.341
CARS	0.819	1.109	0.844	0.970	2.528
RF	0.832	1.069	0.870	0.919	2.668
GLCM	0.892	0.831	0.864	0.915	2.681
Gabor	0.909	0.765	0.866	0.902	2.717
LBP	0.913	0.746	0.864	0.897	2.734
SpeMN	0.872	0.785	0.901	0.819	2.995
SpeSpaMN	0.920	0.689	0.915	0.741	3.311

Table 9
Prediction results of soil carbon content by multiple feature extraction methods in Jiaozhou Bay.

Method	R_c^2	RMSEC	R_p^2	RMSEP	RPD
Full-spectrum	0.629	1.355	0.436	1.544	1.187
UVE	0.473	1.614	0.418	1.485	1.235
SPA	0.510	1.556	0.390	1.501	1.222
GA	0.355	1.785	0.362	1.535	1.194
PCC	0.237	1.942	0.284	1.563	1.173
CARS	0.579	1.443	0.414	1.531	1.198
RF	0.463	1.629	0.391	1.509	1.215
GLCM	0.372	1.761	0.322	1.570	1.168
Gabor	0.405	1.714	0.450	1.426	1.286
LBP	0.422	1.690	0.249	1.715	1.069
SpeMN	0.471	1.533	0.248	1.652	1.110
SpeSpaMN	0.604	1.404	0.346	1.481	1.238

Compared to the full spectral prediction results, R_p^2 of RF was improved by 0.029, RMSEP was decreased by 0.103, and RPD was increased by 0.268. Among the three methods for extracting only spatial features, R_p^2 was generally consistent with the RF, RMSEP was lower than the six methods for extracting only spectral features, and RPD values were higher than the six methods for extracting only spectral features. The best prediction effect was LBP. Compared to the full spectrum prediction results, R_p^2 of LBP was increased by 0.023, RMSEP was decreased by 0.125, and RPD was increased by 0.334. The prediction accuracy of these methods from best to worst was following: SpeSpaMN, SpeMN, LBP, Gabor, GLCM, RF, UVE, SPA, CARS, Full-spectrum, PCC, GA. SpeMN was also a method to extract only spectral features, which is not only superior to the six methods for extracting spectral features, but also superior to the three methods for extracting spatial features. Compared to the full-spectrum prediction results, R_p^2 of SpeMN was improved by 0.06, RMSEP was decreased by 0.203, and RPD was increased by 0.595. SpeSpaMN was the method for extracting spatial and spectral features, with the best prediction effect. Compared to the full-spectrum prediction results, R_p^2 of SpeSpaMN was increased by 0.074, RMSEP decreased was by 0.281, RPD was increased by 0.911. The prediction accuracy of soil carbon content in Aoshan Bay had been significantly improved through SpeSpaMN, and more accurate prediction of soil carbon content could be realized.

In the soil carbon content prediction results by multiple feature extraction methods in Jiaozhou Bay, full-spectrum prediction accuracy was worse. After extracting only spectral features, based on the analysis of RMSEP and RPD, all the spectral feature extraction methods were superior to full-spectrum except for PCC. Based on the analysis of R_p^2, R_p^2 of the full-spectrum was the best. Among the three methods for extracting only spatial features, only prediction results of Gabor was better than those of full-spectrum and six spectral features extraction methods. The prediction accuracy of these methods from best to worst was following: Gabor, SpeSpaMN, UVE, SPA, RF, CARS, GA, Full-spectrum, PCC, GLCM, SpeMN, LBP. The prediction accuracy of SpeMN was not as good as full spectrum. Compared to full-spectrum, SpeSpaMN had a decrease in RMSEP and an increase in RPD, but R_p^2 was lower than full-spectrum. Although some feature extraction methods could improve the prediction accuracy of soil carbon content in Jiaozhou Bay, their improvement was limited, and neither feature extraction method could better predict soil carbon content.

In the three plots, the soil carbon content prediction effects were the best based on SpeSpaMN, followed by SpeMN. Compared with the full-spectrum of HSI, the RPD of the Inland, Aoshan Bay and Jiaozhou Bay samples increased by 47.36 %, 37.96 % and 4.30 %, respectively.

The prediction accuracy of different regression methods were compared to that of SpeMN and SpeSpaMN. The soil samples of three sample plots were divided into calibration set and prediction set in ratio of 2:1 according to the serial number order. After normalizing the hyperspectral data, the soil carbon content models in the three plots were established by using SpeMN and SpeSpaMN, and compared with the soil carbon content prediction results established by principal component regression (PCR) (Ribeiro et al., 2021), partial least square regression (PLSR) (Wang et al., 2022), least squares support vector machines (LSSVM) (Li et al., 2019), back propagation neural network (BPNN) (Li et al., 2019), long short-term memory (LSTM) (Wang et al., 2022), and Convolutional Neural Networks (CNN) (Wang et al., 2022). Prediction results of soil carbon content by multiple modeling methods in three sample plots were shown in the Table 10.

In the soil carbon content prediction results by multiple modeling methods in Inland, LSTM had the best prediction results among PCR, PLSR, LSSVM, BPNN, LSTM and CNN, which can predict soil carbon content. Compared to the six methods, the proposed SpeMN method was better than the LSTM method. By using the SpeSpaMN method, the prediction results of soil carbon content were significantly improved, which was better than other methods and could predict the soil carbon content more accurately. Extracting hyperspectral feature information by SpeMN and SpeSpaMN could improve the prediction accuracy of Inland soil carbon content to varying degrees.

In the soil carbon content prediction results by multiple modeling methods in Aoshan Bay, PCR, PLSR, LSSVM, LSTM and CNN could better predict the soil carbon content except for BPNN. The prediction effect of the SpeMN model was further improved compared with the other five methods. SpeSpaMN had the best prediction effect. SpeMN could improve the prediction results of soil carbon content model in Aoshan Bay, and the SpeSpaMN improve prediction effect significantly.

In the soil carbon content prediction results by multiple modeling methods in Jiaozhou Bay, LSSVM had the best prediction results. The prediction effect of SpeMN model was only superior to PCR, BPNN and CNN. SpeSpaMN was superior to PCR, PLSR, BPNN, LSTM and CNN, but slightly inferior to LSSVM. In the soil carbon content model of Jiaozhou Bay, neither model could predict the soil carbon content well. The prediction effects were slightly improved by using the SpeMN and

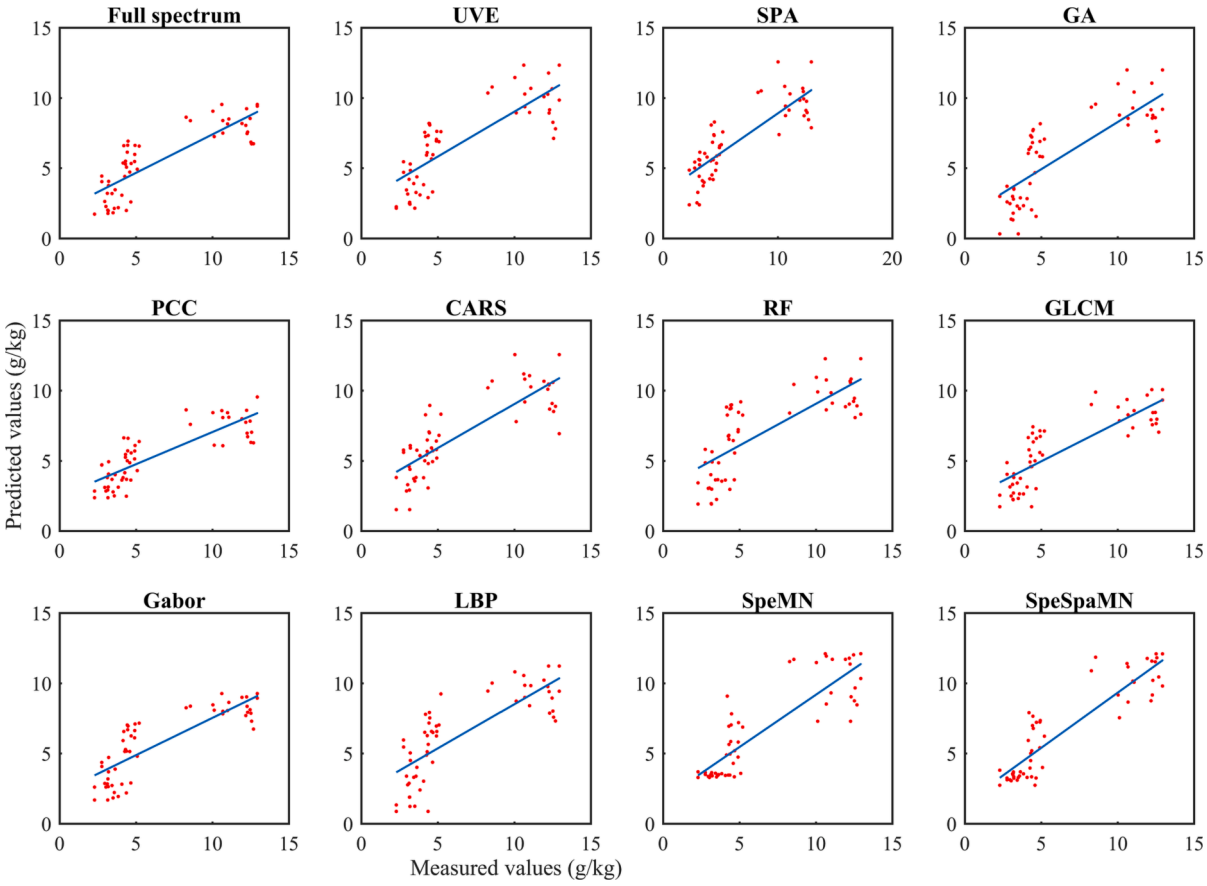


Fig. 11. Fitting results of soil carbon content measured values and predicted values by multiple feature extraction methods in Inland.

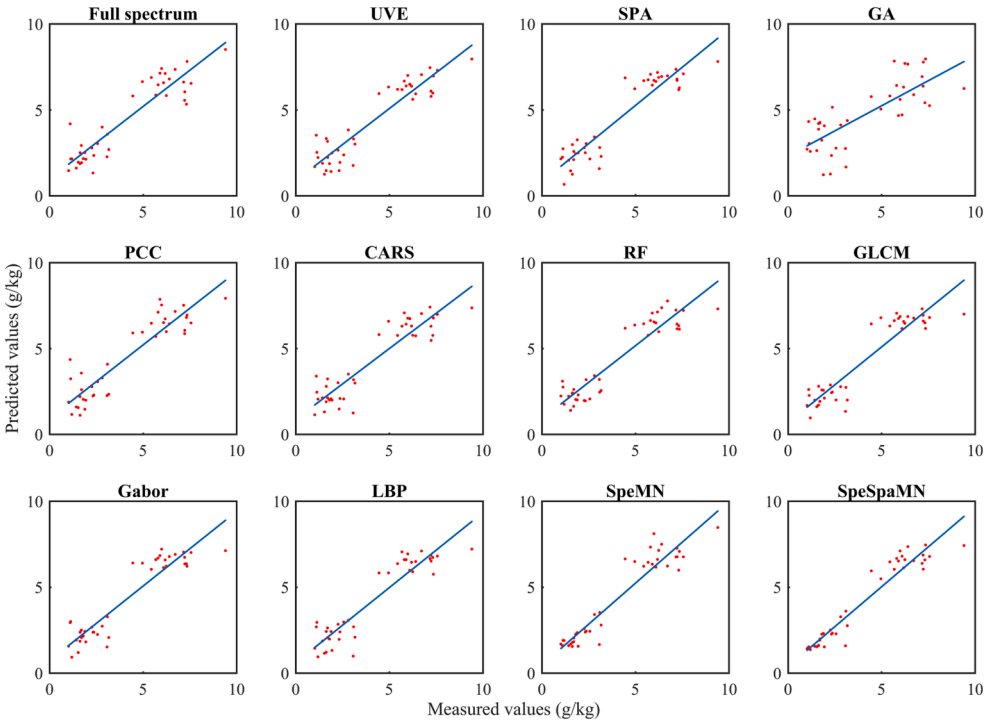


Fig. 12. Fitting results of soil carbon content measured values and predicted values by multiple feature extraction methods in Aoshan Bay.

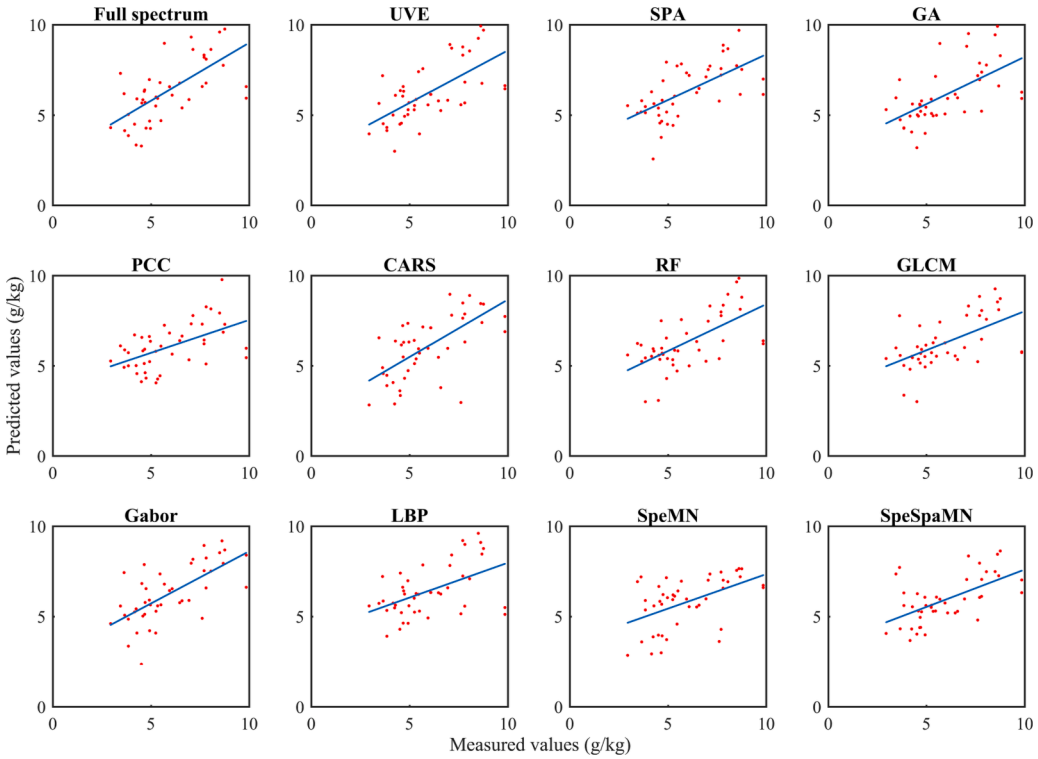


Fig. 13. Fitting results of soil carbon content measured values and predicted values by multiple feature extraction methods in Jiaozhou Bay.

Table 10
Prediction results of soil carbon content by multiple modeling methods in three sample plots.

Method	Inland			Aoshan Bay			Jiaozhou Bay		
	R_p^2	RMSEP	RPD	R_p^2	RMSEP	RPD	R_p^2	RMSEP	RPD
PCR	0.618	3.257	1.139	0.846	1.009	2.430	0.033	2.222	0.825
PLSR	0.669	2.419	1.533	0.841	1.022	2.400	0.436	1.544	1.187
LSSVM	0.619	2.379	1.560	0.864	0.953	2.573	0.420	1.443	1.271
BPNN	0.591	2.358	1.573	0.730	1.263	1.941	0.233	1.752	1.047
LSTM	0.715	2.098	1.768	0.873	0.928	2.643	0.410	1.537	1.193
CNN	0.619	2.270	1.634	0.897	0.821	2.987	0.247	1.980	0.926
SpeMN	0.728	1.918	1.934	0.901	0.819	2.995	0.248	1.652	1.110
SpeSpaMN	0.801	1.642	2.259	0.915	0.741	3.311	0.346	1.481	1.238

SpeSpaMN.

By analyzing the prediction results of soil carbon content using multiple models in three sample plots, SpeMN and SpeSpaMN could improve the prediction effects to different degrees for models with general and good prediction results established by existing methods. Taking the Inland and Aoshan Bay sample plots as examples, the SpeMN improved the prediction effects, and SpeSpaMN with adding spatial information feature extraction had the best prediction effect. For the models with worse prediction results established by existing methods, such as the Jiaozhou Bay sample plot, SpeMN and SpeSpaMN could improve the prediction effects compared to some models, but could not achieve the optimal prediction effect. The overall prediction results of Jiaozhou Bay were not ideal. The reason might be that the carbon content values of the Jiaozhou Bay samples were relatively close (the SD value was the smallest in the three sample plots, table 1), and the spectral differentiation of each sample was small, so the prediction effect based on HSI was worse. Another reason might be that the training set dividing according to the serial number order did not have the representativeness of all samples, which resulted in the worse prediction effect of each method, especially the proposed SpeMN and SpeSpaMN methods to achieve the optimal prediction results. The subsequent experiments would use different methods of dividing training sets t.

o further validate the prediction results of each method.

To verify the prediction effects of SpeMN and SpeSpaMN for dividing the calibration set and prediction sets in different ways, the KS method and SPXY method were used to divide calibration set and prediction sets in a ratio of 2:1, respectively. Soil carbon content models were established by SpeMN and SpeSpaMN, which compared with the PCR, PLSR, LSSVM, BPNN, LSTM and CNN models. Prediction results of soil carbon content by different calibration set and prediction set division methods

Table 11
Prediction results of soil carbon content by different calibration set and prediction set division methods in Inland.

Method	KS			SPXY		
	R_p^2	RMSEP	RPD	R_p^2	RMSEP	RPD
PCR	0.406	2.734	1.304	0.440	2.187	1.324
PLSR	0.596	2.278	1.565	0.582	1.907	1.519
LSSVM	0.560	2.347	1.519	0.613	1.848	1.568
BPNN	0.595	2.261	1.577	0.603	1.859	1.558
LSTM	0.684	2.284	1.684	0.690	1.713	1.748
CNN	0.710	2.213	1.738	0.849	1.319	2.271
SpeMN	0.782	2.154	1.785	0.881	1.162	2.578
SpeSpaMN	0.793	1.952	1.970	0.882	1.049	2.854

Table 12
Prediction results of soil carbon content by different calibration set and prediction set division methods in Aoshan Bay.

Method	KS			SPXY		
	R_p^2	RMSEP	RPD	R_p^2	RMSEP	RPD
PCR	0.533	1.688	1.423	0.922	0.752	3.245
PLSR	0.566	1.580	1.520	0.909	0.733	3.331
LSSVM	0.533	1.656	1.450	0.934	0.727	3.358
BPNN	0.835	1.105	2.315	0.940	0.653	3.740
LSTM	0.581	1.720	1.487	0.890	0.812	3.010
CNN	0.889	0.983	2.607	0.926	0.756	3.342
SpeMN	0.909	0.789	3.243	0.973	0.522	4.682
SpeSpaMN	0.912	0.762	3.357	0.973	0.475	5.141

Table 13
Prediction results of soil carbon content by different calibration set and prediction set division methods in Jiaozhou Bay.

Method	KS			SPXY		
	R_p^2	RMSEP	RPD	R_p^2	RMSEP	RPD
PCR	0.099	1.662	0.960	0.352	1.294	1.226
PLSR	0.234	1.485	1.074	0.506	1.200	1.322
LSSVM	0.243	1.414	1.128	0.525	1.190	1.334
BPNN	0.242	1.681	1.070	0.310	1.423	1.115
LSTM	0.298	2.238	0.804	0.678	1.045	1.634
CNN	0.283	1.788	1.006	0.689	1.350	1.264
SpeMN	0.359	1.461	1.232	0.722	0.932	1.832
SpeSpaMN	0.389	1.436	1.253	0.731	0.884	1.930

in Inland, Aoshan Bay and Jiaozhou Bay were shown in the Table 11, Table 12 and Table 13, respectively.

In the prediction results of KS and SPXY division methods, the prediction results of SPXY were better than those of KS in three sample plots. This was because SPXY method considered the carbon content of each sample when dividing the calibration set and the prediction set, and the divided calibration set was more representative in the distribution, so the prediction effect was better. Taking inland samples as an example, it could be seen from the principal component spatial distribution diagram in two division methods (Fig. 14) that the calibration set and prediction set divided by SPXY had a uniform sample distribution, which was more conducive to the establishment of the model. In the soil carbon content prediction results by different calibration set and prediction set division methods, the prediction results of SpeMN and SpeSpaMN in each sample plot were superior to the other six methods, and SpeSpaMN had the best prediction results. No matter which division method was used, SpeMN and SpeSpaMN showed the best prediction effects, verifying the accuracy and stability of SpeMN and SpeSpaMN methods.

In this paper, the prediction effects in soil carbon prediction were improved by introducing the spatial information. Future work will analyze the contribution of hyperspectral spectral and spatial information in the prediction of soil carbon content. The spatial and spectral information of HSI are fused according to the difference in contribution, which to further improve the prediction effects of soil carbon content. Due to the difficulty in obtaining soil samples, future work will study the HSI soil carbon content prediction under small samples through deep learning. Due to the complexity of the proposed algorithm structure, it takes longer computation time compared to machine learning methods, future work will study on the lightweight networks in HSI soil carbon content prediction to reduce computation time.

4. Conclusions

The hyperspectral images of soil samples from Inland, Aoshan Bay, and Jiaozhou Bay were taken as examples, multi-scale spectral feature network (SpeMN) and multi-scale spectral and spatial feature network (SpeSpaMN) were proposed to establish soil carbon content models for three sample plots. SpeMN used the spectral vector of multiple scales as input to obtain spectral features of multiple scales, and realized the complementarity among features at different scales. SpeSpaMN fused multi-scale spatial feature networks based on SpeMN, which extracted both multi-scale spectral feature information and different scales of spatial context features in HSI. Among the three plots, soil carbon content prediction was the best based on SpeSpaMN, followed by SpeMN. The RPD of SpeSpaMN in Inland, Aoshan Bay and Jiaozhou Bay samples were 2.259, 3.311 and 1.238, respectively, which were 47.36 %, 37.96 % and 4.30 % higher than the full-spectrum of hyperspectral images. The accuracy and stability of SpeMN and SpeSpaMN were further verified by comparing different feature extraction methods, different regression methods, different calibration set and prediction set division methods. This paper can effectively solve the problem of the deep fusion of spatial and spectral features in the soil carbon content prediction by hyperspectral image, so as to improve the accuracy and stability of soil carbon content prediction, promote the application and development of soil carbon content prediction in hyperspectral image, and provide technical support for the study of carbon cycle and carbon sink.

This work was supported by Natural Science Foundation of Shandong Province, grant number is ZR2021QF028, ZR2021MD093 and

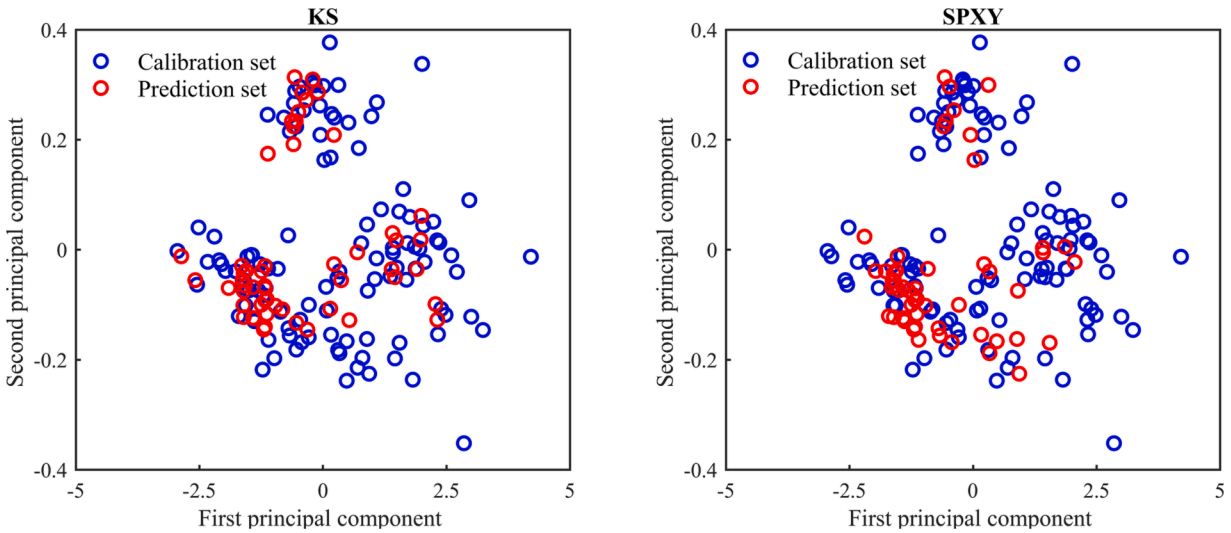


Fig. 14. Principal components spatial distribution diagram by KS and SPXY dividing the calibration set and prediction set at Inland.

ZR2021MD103, and the National Natural Science Foundation of China, grant number is 32171578 and 42307483, Natural Science Foundation of Qingdao City, grant number is 23–2-1–211-zyyd-jch, Qilu University of Technology (Shandong Academy of Sciences), grant number is 2023HYZX01 and 2023PX037, Shandong Province Key Research and Development Program, grant number is 2023ZLYS01.

CRediT authorship contribution statement

Xueying Li: Conceptualization, Funding acquisition, Methodology, Writing – original draft. **Zongmin Li:** Formal analysis, Funding acquisition, Supervision. **Huimin Qiu:** Formal analysis, Validation, Funding acquisition. **Guangyuan Chen:** Formal analysis, Validation. **Pingping Fan:** Funding acquisition, Writing – review & editing. **Yan Liu:** Funding acquisition, 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.

Data availability

The data that has been used is confidential.

References

Bao, Y., Meng, X., Ustin, S., Wang, X., Zhang, X., Liu, H., Tang, H., 2020. Vis-SWIR spectral prediction model for soil organic matter with different grouping strategies. Catena 195, 104703.
Brown, D.J., 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. Geoderma 140 (4), 444–453.
Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2018. Evaluating the capability of the sentinel 2 data for soil organic carbon prediction in croplands. ISPRS J. Photogramm. Remote Sens. 147, 267–282.
Chu, R., Richard, N., Chatoux, H., Fernandez-Maloigne, C., Hardeberg, J., 2021. Hyperspectral texture metrology based on joint probability of spectral and spatial distribution. IEEE Trans. Image Process. 30, 4341–4356.
Dristi, D., Manoranjan, P., Manzur, M., Shyh, W.T., Leigh, S., 2022. Soil moisture, organic carbon, and nitrogen content prediction with hyperspectral data using regression models. Sensors 22, 7998.
Emile, N., Dinh, H.T.M., Nicolas, B., Dominique, C., Laure, H., 2018. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for camargue. France. Remote Sensing 10 (8), 1217.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hou, T., Sun, W., Chen, C., Yang, G., Meng, X., Peng, J., 2022. Marine floating raft aquaculture extraction of hyperspectral remote sensing images based decision tree algorithm. Int. J. Appl. Earth Obs. Geoinf. 111, 102846.
Hu, W., Li, H., Pan, L., Li, W., Tao, R., Du, Q., 2020. Spatial-spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 58 (6), 4237–4250.
Hu, W., Li, H., Deng, Y., Sun, X., Du, Q., Plaza, A., 2021. Lightweight tensor attention-driven ConvLSTM neural network for hyperspectral image classification. IEEE J. Sel. Top. Signal Process. 15 (3), 734–745.
Imani, M., Ghassemian, H., 2020. An overview on spectral and spatial information fusion for hyperspectral image classification: current trends and challenges. Information Fusion 59, 59–83.
Jaber, S.M., Lant, C.L., Al-Qinna, M.I., 2011. Estimating spatial variations in soil organic carbon using satellite hyperspectral data and map algebra. Int. J. Remote Sens. 32 (18), 5077–5103.
Lecun, Y., Bottou, L., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Li, W., Chen, C., Su, H., Du, Q., 2015. Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Trans. Geosci. Remote Sens. 53 (7), 3681–3693.
Li, X., Fan, P., Liu, Y., Hou, G., Wang, Q., Lv, M., 2019. Prediction results of different modeling methods in soil nutrient concentrations based on spectral technology. J. Appl. Spectrosc. 86 (4), 765–770.
Li, S., Hao, W., Wan, D., Zhu, J., 2011. An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. Knowl.-Based Syst. 24 (1), 40–48.
Li, X., Li, Z., Qiu, H., Chen, G., Fan, P., 2023. Soil carbon content prediction using multi-source data feature fusion of deep learning based on spectral and hyperspectral images. Chemosphere 336, 139161.
Liu, K., Sun, W., Shao, Y., Liu, W., Yang, G., Meng, X., Peng, J., Mao, D., Ren, K., 2022. Mapping coastal wetlands using transformer in transformer deep network on China ZY1-02D hyperspectral satellite images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 3891–3903.
Liu, Q., Zhou, F., Hang, R., Yuan, X., 2017. Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification. Remote Sens. (Basel) 9 (12), 1330.
Meng, X., Bao, Y., Ye, Q., Liu, H., Zhang, X., Tang, H., Zhang, X., 2021. Soil organic matter prediction model with satellite hyperspectral image based on optimized denoising method. Remote Sens. (Basel) 13, 2273.
Michael, B.F., Helen, M.W., Iman, T., Catherine, M.Y., Peter, K.D., Shahla, H.B., 2023. Rapid assessment of soil carbon and nutrients following application of organic amendments. Catena 223, 106928.
Mou, L., Ghamisi, P., Zhu, X., 2017. Deep recurrent neural networks for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 55 (7), 3639–3655.
Petri, P., Markku, L., Niklas, S., Jesse, H., Ilja, V., Matti, R., Janne, H., Mika, S., Kristiina, K., Arto, K., 2023. Tropical altitudinal gradient soil organic carbon and nitrogen estimation using specim IQ portable imaging spectrometer. Sci. Total Environ. 833, 163677.
Ribeiro, S.G., Teixeira, A.D., de Oliveira, M.R.R., Costa, M.C.G., Araujo, I.C.D., Moreira, L.C.J., Lopes, F.B., 2021. Soil organic carbon content prediction using soil-reflected spectra: a comparison of two regression methods. Remote Sens. (Basel) 13, 4752.
Shi, X., Chen, Z., Wang, H., Yeung, D.Y., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems 28.
Singh, S., Kasana, S.S., 2019. Estimation of soil properties from the EU spectral library using long short-term memory networks. Geoderma Reg. 18, e00233.
Sun, W., Yang, G., Peng, J., Meng, X., He, K., Li, W., Li, H., Du, Q., 2022. A multiscale spectral features graph fusion method for hyperspectral band selection. IEEE Trans. Geosci. Remote Sens. 60, 1–12.
Vohland, M., Ludwig, M., Thiele-Bruhn, S., 2017. Quantification of soil properties with hyperspectral data: selecting spectral variables with different methods to improve accuracies and analyze prediction mechanism. Remote Sens. (Basel) 9, 1103.
Wang, D., Du, B., Zhang, L., Xu, Y., 2021a. Adaptive spectral-spatial multiscale contextual feature extraction for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 59 (3), 2461–2477.
Wang, S., Guan, K., Zhang, C., Lee, D., Margenot, A.J., Ge, Y., Peng, J., Zhou, W., Zhou, Q., Huang, Y., 2022. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: assessing potential of airborne and spaceborne optical soil sensing. Remote Sens. Environ. 271, 112914.
Wang, X., Han, J., Wang, X., Yao, H., Zhang, L., 2021c. Estimating soil organic matter content using Sentinel-2 imagery by machine learning in Shanghai. IEEE Access 9, 78215–78225.
Wang, K., Qi, Y., Guo, W., Zhang, J., Chang, Q., 2021b. Retrieval and mapping of soil organic carbon using sentinel-2A spectral images from bare cropland in autumn. Remote Sens. (Basel) 13, 1072.
Wei, X., He, J., Zheng, S., Ye, D., 2020. Modeling for SSC and firmness detection of persimmon based on NIR hyperspectral imaging by sample partitioning and variables selection. Infrared Phys. Technol. 105.
Wei, L., Wang, K., Lu, Q., Liang, Y., Li, H., Wang, Z., Wang, R., Cao, L., 2021. Crops fine classification in airborne hyperspectral imagery based on multi-feature fusion and deep learning. Remote Sens. (Basel) 13 (15), 2917.
Weng, S., Tang, P., Yuan, H., Guo, B., Yu, S., Huang, L., Xu, C., 2020a. Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion. Spectrochim. Acta A Mol. Biomol. Spectrosc. 234, 118237.
Weng, S., Yu, S., Guo, B., Tang, P., Liang, D., 2020b. Non-destructive detection of strawberry quality using multi-features of hyperspectral imaging and multivariate methods. Sensors 20 (11), 3074.
Xu, L., Hong, Y., Wei, Y., Guo, L., Shi, T., Liu, Y., Jiang, Q., Fei, T., Liu, Y., Mouazen, A., Chen, Y., 2020a. Estimation of organic carbon in anthropogenic soil by VIS-NIR spectroscopy: effect of variable selection. Remote Sens. (Basel) 12 (20), 3394.
Xu, S., Liu, S., Wang, H., Chen, W., Zhang, F., Xiao, Z., 2020b. A hyperspectral image classification approach based on feature fusion and multi-layered gradient boosting decision trees. Entropy 23 (1), 20.
Yang, G., Huang, K., Sun, W., Meng, X., Mao, D., Ge, Y., 2022. Enhanced mangrove vegetation index based on hyperspectral images for mapping mangrove. ISPRS J. Photogramm. Remote Sens. 289, 236–254.
Yang, J., Wang, X., Wang, R., Wang, H., 2020. Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis-NIR spectroscopy. Geoderma 380, 114616.
Zhang, R., Xie, H., Cai, S., Hu, Y., Liu, G., Hong, W., Tian, Z., 2020. Transfer-learning-based raman spectra identification. J. Raman Spectrosc. 51 (1), 176–186.