

Progressive Layered Extraction Network Based on Correlation Sharing for Multi-target Prediction of Soil Nutrients

Tielong SU¹, Xuesong TIAN², Zhengguang CHEN^{1*}

1. College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China; 2. Daqing Oilfield Shale Oil Exploration and Development Headquarters, Daqing 163319, China

Abstract With breakthroughs in data processing and pattern recognition through deep learning technologies, the use of advanced algorithmic models for analyzing and interpreting soil spectral information has provided an efficient and economical method for soil quality assessment. However, traditional single-output networks exhibit limitations in the prediction process, particularly in their inability to fully utilize the correlations among various elements. As a result, single-output networks tend to be optimized for a single task, neglecting the interrelationships among different soil elements, which limits prediction accuracy and model generalizability. To overcome this limitation, in this study, a multi-task learning architecture with a progressive extraction network was implemented for the simultaneous prediction of multiple indicators in soil, including nitrogen (N), organic carbon (OC), calcium carbonate (CaCO_3), cation exchange capacity (CEC), and pH. Furthermore, while incorporating the Pearson correlation coefficient, convolutional neural networks, long short-term memory networks and attention mechanisms were combined to extract local abstract features from the original spectra, thereby further improving the model. This architecture is referred to as the Relevance-sharing Progressive Layered Extraction Network. The model employs an adaptive joint loss optimization method to update the weights of individual task losses in the multi-task learning training process.

Key words Near-infrared spectroscopy; Progressive extraction network; Multi-task learning; Convolutional neural network; Long short-term memory network; Attention mechanism

DOI:10.19759/j.cnki.2164-4993.2025.05.007

Materials and Methods

Data source

The dataset consists of soil spectral samples collected from the Land Use/Cover Area frame Statistical Survey (LUCAS) conducted by Eurostat between 2008 and 2012. Sampling points were distributed in 23 member states of the European Union. The main sampling areas included cropland, forest land, and grassland. Soil spectral measurements were obtained using a Foss XDS spectrometer, with a wavelength range of 400–2 500 nm, a resolution of 0.5 nm, and a total of 4 200 wavelength points^[1]. The spectral data of all samples are shown in Fig. 1(a).

The LUCAS dataset encompasses a variety of physical, chemical, and biological characteristics of soil. In terms of physico-chemical attributes, the dataset provides detailed records of the following indicators: pH, organic carbon (OC) content, nutrient concentrations including nitrogen and calcium carbonate which are essential for plant growth, as well as cation exchange capacity, which contains extensive soil information.

Spectral data processing

This study was conducted to predict multiple soil characteristics, specifically focusing on five characteristics: pH (pH in H_2O), organic carbon content (OC), total nitrogen content (N), cation exchange capacity (CEC), and calcium carbonate (CaCO_3). A total of 19 036 samples were selected for the experiment and randomly divided into a training set (15 228) and a test

set (3 808) in a 4 : 1 ratio. Due to the significant baseline drift in the spectra (Fig. 1(a)), in this study, Standard Normal Variate transformation (SNV)^[2] was employed to eliminate baseline drift in the spectra. The soil spectral curves after SNV processing are shown in Fig. 1(b).

MTL Neural Network Prediction

Progressive layered extraction network

The Progressive Layered Extraction (PLE) network^[3] is a novel architecture proposed for the field of Multi-Task Learning (MTL), designed to address common issues in traditional MTL such as negative transfer and the seesaw phenomenon. As shown in Fig. 2(a), the PLE model explicitly separates the feature extraction layer into two components: Shared Experts and Task-Specific Experts. It introduces a gating network to balance information sharing and task-specific information. In the figure, Expert A and Expert B represent the respective expert systems for Task A and Task B, while Shared denotes the shared expert system.

The expert system is a specific deep learning network tailored for task prediction. Therefore, for any task k , e_k represents the feature matrix extracted by its expert network. $S_k(x)$ concatenates the features extracted by the Shared Experts and the Task-Specific Experts of task k , forming a selected matrix composed of the outputs from the Shared Experts and the Experts-K for task k .

$$S_k^k(x) = [E_{(k,1)}^T, E_{(k,2)}^T, \dots, E_{(k,n)}^T, E_{(s,1)}^T, E_{(s,2)}^T, \dots, E_{(s,n)}^T]^T \quad (1)$$

The structure of the gating network is typically a single-layer feedforward network, as shown in the lower right part of Fig. 2(a). The calculation method for the Gate is as follows:

$$w^k(x) = \text{softmax}(w_g^k x) \quad (2)$$

In the equation, x represents the input; $w^k(x)$ denotes the

Received: July 15, 2025 Accepted: September 19, 2025

Tielong SU (2000–), male, P. R. China, devoted to research about deep learning and near infrared spectroscopy.

* Corresponding author.

weighting function for the k^{th} task; and $\text{softmax}()$ is the activation function, which assigns appropriate weight vectors to the outputs of the expert networks and shared network; and w_g^k represents the weight matrix in the gating network. The output of the Gate can be expressed as:

$$g^k(x) = w^k(x) S^k(x) \quad (3)$$

The output of the gating unit is fed into the tower network, where the data is processed to obtain the final prediction result. The calculation method of the tower network is as follows:

$$y^k(x) = t^k(g^k(x)) \quad (4)$$

In the equation, t^k represents the computation of the tower network for task k , while $y^k(x)$ denotes the predicted output for task k . To accommodate the preference for abstract features among different tasks, this experiment focused on predictive task elements with strong correlations. As shown in Fig. 2(b), based on the PLE architecture, when Task A and Task B exhibit high

correlation, the outputs from the first-layer experts of the two tasks are partially shared. After a weighted summation through the gating network, they are fed into the next layer of experts for further feature extraction. This architecture is referred to as the Relevance-sharing Progressive Layered Extraction network (R-PLE). The input to the gating network in the R-PLE model can be expressed as:

$$S^k(x) = [E_{(k,1)}^T, E_{(k,2)}^T, \dots, E_{(k,n)}^T, E_{(s,1)}^{(s,1)}, E_{(s,2)}^T, \dots, E_{(s,n)}^T, E_{(R,2)}^T, \dots, E_{(R,n)}^T] \quad (5)$$

$E_{(R,n)}^T$ represents the expert layer output of tasks highly correlated with task k . After integrating partial features, the input and output feature dimensions remain unchanged. The optimized model more effectively leverages inter-task correlations and shares more critical features, thereby further enhancing the gating network's filtering capability.

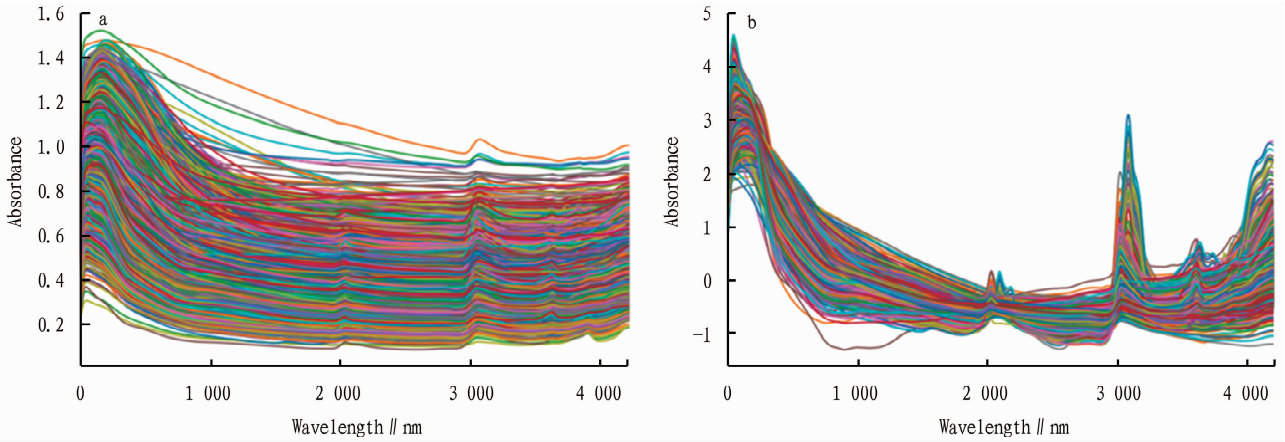


Fig. 1 Dimensionality reducing near infrared (NIR) spectra of LUCAS dataset (a) and SNV-preprocessed NIR spectra (b)

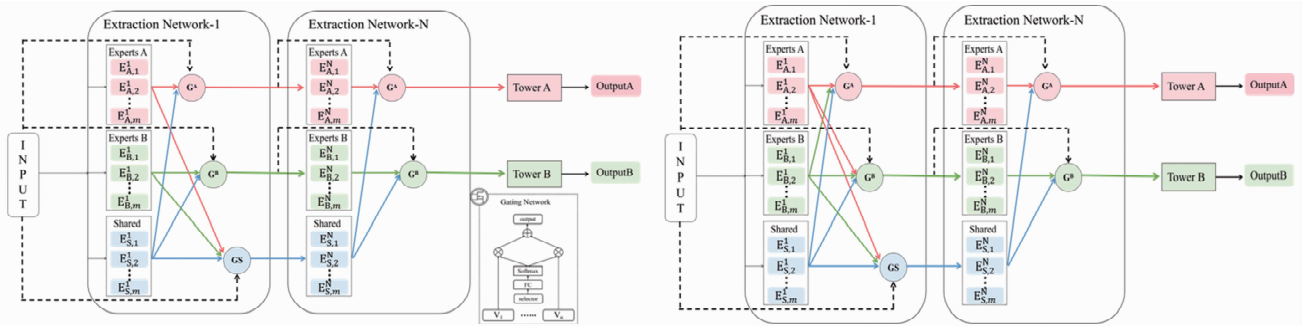


Fig. 2 Forecast model of PLE (a) and forecast model of R-PLE (b)

Model Design

Model structure

Based on the Progressive Layered Extraction (PLE) network and the correlations between different tasks, a multi-task learning network called R-PLE with dedicated expert layers was proposed in this study. This structure is referred to as RCLA-PLE in this study. Fig. 3 illustrates the overall architecture of the system and its detailed composition. The right half provides an overview of the entire system, showcasing the main components and their interrela-

tionships, offering a macro-level perspective. The left half (P1-P3) presents detailed cross-sectional views of this overall architecture, delving into the internal structure and operational principles of each key component.

The expert layer is designed to adaptively extract input feature information. The model adopts a two-layer feature extraction module (Extraction Network_1 to Extraction Network_2), where the network input is one-dimensional raw soil Vis-NIR data, and the output consists of multiple properties of the target soil. In processing

one-dimensional sequential data, deep convolutional networks (CNN)^[4] can progressively extract features from low-level to high-level by stacking multiple convolutional layers, thereby enhancing the model's expressive capability. Based on this, the first-layer feature extraction module in this study adopts a structure composed of seven layers of stacked one-dimensional deep convolutional blocks (Conv1D block_1 to Conv1D block_7) to extract different local features from the input spectrum. Each layer of convolutional blocks includes convolution operations, activation functions, and pooling operations. The design of the second feature extraction layer (Extraction Network-2) aims to further enhance the model's performance. To capture the potential relationships between spectral data of different wavelength bands, a Long Short-Term Memory (LSTM) network^[5] is introduced. Since convolutional operations effectively extract local patterns and spatial information, the second expert layer adopts a single convolutional structure and multiple LSTM layers for further processing. The LSTM module consists of two LSTM layers. The first LSTM layer takes the output feature sequence extracted by the CNN module as input and generates features for each time step. The number of hidden memory units in the first LSTM layer is set to 32, which equals the dimensionality of the output features from this layer. The number of hidden memory units in the second layer is set to 8, enabling the extraction of higher-level local and abstract features from the output features of the first layer. Finally, the output of the second LSTM layer at the last time step is passed to the Tower layer for the final regression prediction.

Joint loss optimization

In conventional multi-task models, the data distribution and importance of different tasks often vary. Due to significant differences in loss scales among different task outputs, a weighted sum of each task^[6] is selected as the minimized adaptation loss for multi-task learning to prevent the overall model loss from being dominated by a single task. It can generally be expressed as:

$$L_{MTL} = \sum_i W_i L_i \quad (6)$$

$$L(W, \sigma_1, \sigma_2, \dots, \sigma_5) = \frac{1}{2} \sum_{i=1}^5 \frac{1}{\sigma_i^2} L_i(W) + \sum_{i=1}^5 \log \sigma_i \quad (7)$$

In this study, uncertainty optimization was employed for the multi-task learning training process. The final loss function can be expressed as Equation (7). Through maximization of the Gaussian likelihood, the task uncertainty σ_i is transformed into an adaptive loss weight of $1/(2\sigma_i^2)$. During training, both the model parameters W and the uncertainty parameters σ_i are optimized simultaneously. Tasks with higher uncertainty automatically receive smaller weights, thereby achieving automatic balancing of

the multi-task loss.

Conclusions and Discussion

As shown in Fig. 4, the scatter points of the RCLA-PLE model are tightly clustered near the diagonal line, particularly for pH predictions, which demonstrate lower errors and better fitting performance. In the OC scatter plot, while the median predictions are relatively dispersed, the diagonal region shows a higher concentration of points with denser and more compact distribution, indicating smaller and more stable prediction errors of the model. Furthermore, in the CaCO_3 predictions, the scatter points of the RCLA-PLE model are evenly distributed along the diagonal line, indicating that its predictions are closer to the true values. The model performs optimally in N predictions, demonstrating that RCLA-PLE can effectively capture the spatial distribution and temporal dynamics of nitrogen content.

Model comparison

To further evaluate the performance of different modeling approaches, the results of the RCLA-PLE model proposed in this paper were compared with the single-task convolutional neural network (CNN) used by Padarian *et al.*^[7], the multi-task convolutional network (Multi-CNN) proposed by Padarian *et al.* in the same year^[7], the long short-term memory (LSTM) network used by Singh^[8], and the combined CNN and GRU model (CCNVR) used by Yang^[9]. The model's performance was evaluated using the following three metrics: Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Error (MAE). The complete set of results is presented in Table 2. Due to the correlations among different soil properties, Padarian *et al.* found that neural networks can improve model performance when these correlations are considered. In this study, we leveraged data correlations to optimize the model's fitting capability, and focused not only on single-element output methods but also on multi-attribute output approaches, further validating the effectiveness of the proposed model. Moreover, previous studies applied different dataset divisions and employed distinct preprocessing methods (*e.g.*, the 2D spectral maps in Padarian *et al.* and the multi-spectral preprocessing in Tsakiridis *et al.* (2020)^[10]). The results demonstrate that RCLA-PLE can accurately predict each soil property. Except for CEC, where it slightly underperforms compared with the Long Short-Term Memory (LSTM) model used by Simranjit *et al.*, it surpasses all previous single-task and multi-task models in all other aspects.

Table 2 Comparison of proposed framework with previous works

Soil properties	Proposed			Singh <i>et al.</i> [8] (LSTM)			Padarian <i>et al.</i> [7] (CNN)		Padarian <i>et al.</i> [7] (Multi-CNN)		Yang ^[9] (CNN-GRU)	
	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	RMSE	R^2	RMSE	R^2
N	0.91	0.92	0.53	1.15	0.91	0.64	1.54	0.83	1.06	0.60	0.45	0.70
OC	0.67	0.94	9.245 4	23.25	0.94	11.24	32.14	0.88	16.82	0.69	6.40	0.73
CEC	5.85	0.75	3.987 1	6.75	0.77	3.89	8.58	0.66	6.51	0.63	3.30	0.73
pH	0.38	0.91	0.275 3	0.42	0.90	0.32	0.54	0.87	0.53	0.84	0.35	0.86

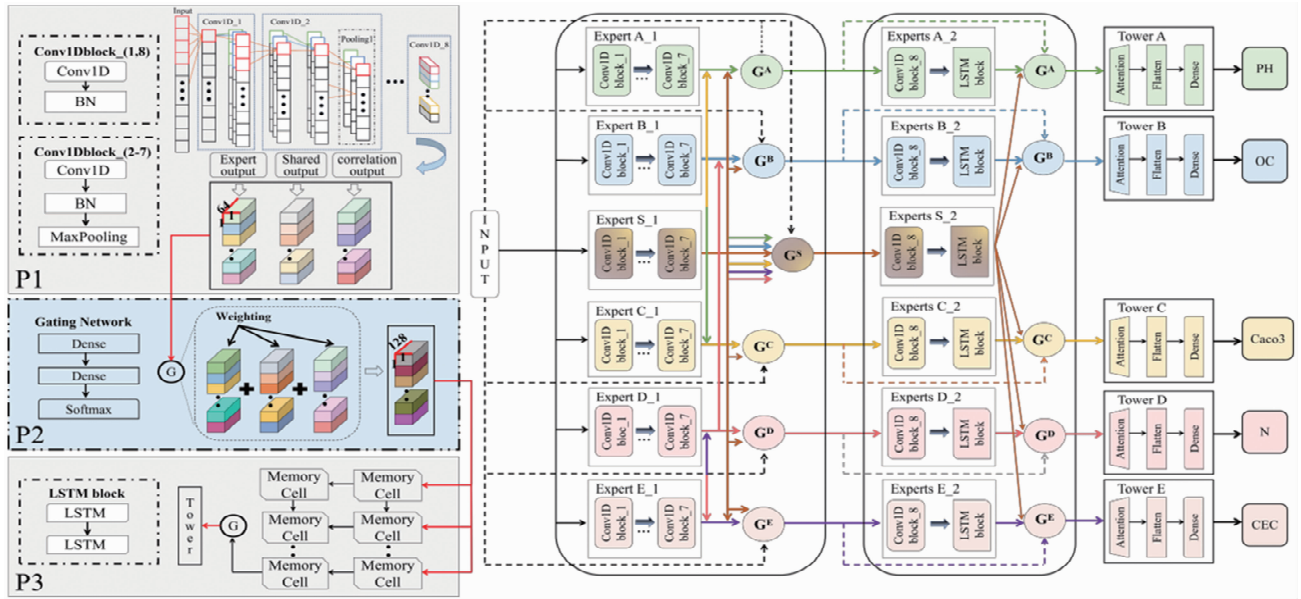
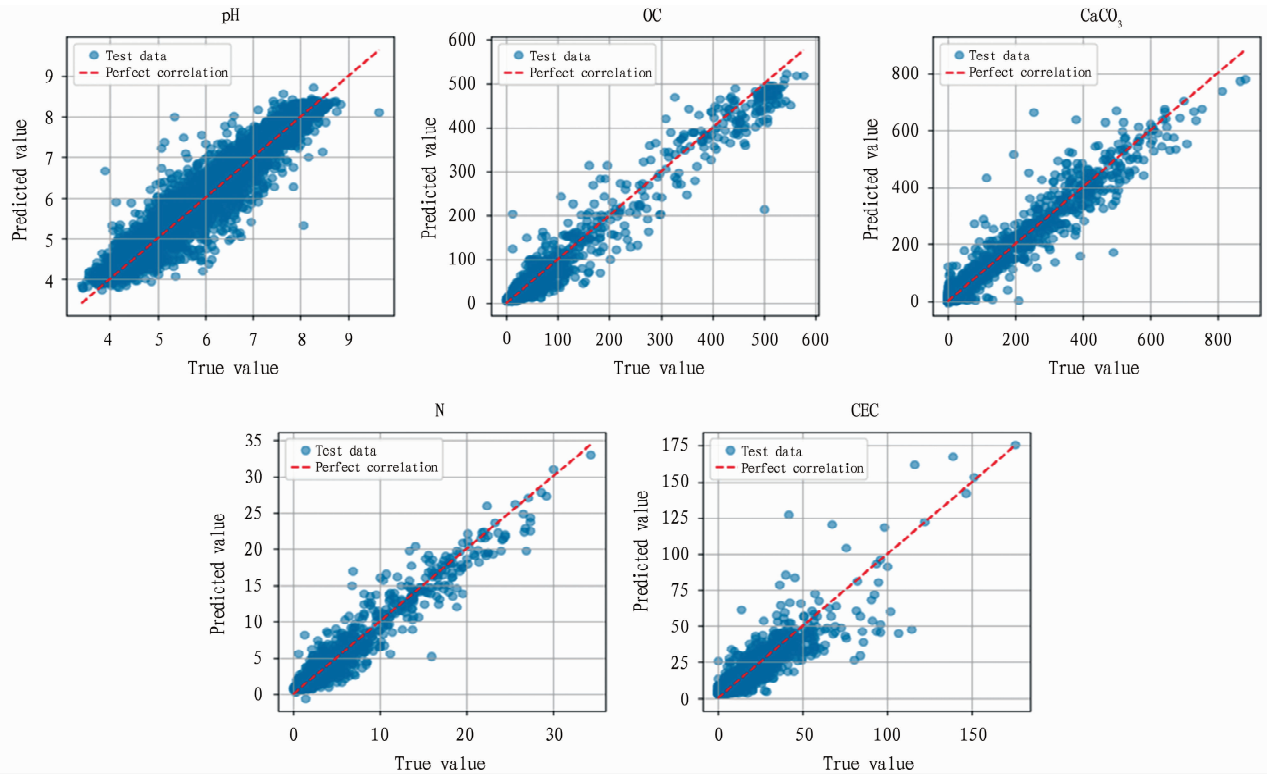


Fig. 3 The overall architecture of the proposed RCLA-PLE model



The red dashed line is the 1 : 1 line.

Fig. 4 RCLA-PLE scatter plot of measured and predicted values for the five soil properties

Conclusions

In this study, correlation coefficient algorithms were integrated with the advantages of multi-task learning to propose a novel RCLA-PLE algorithm. Initially, local features between different elements are shared based on the strength of their correlations. Subsequently, dedicated expert layers composed of CNN, LSTM, and Attention networks are employed to capture intricate relationships

among these features. Finally, the features are combined according to their correlation strengths, resulting in more accurate prediction outcomes. Through the construction of multi-task learning models using the near-infrared spectral dataset from the LUCAS database, the validity of the conclusion that multi-task learning outperforms single-task learning was verified. The results demonstrated that the

(Continued on page 41)

demonstrate MCM-Net’s superior boundary smoothness and object integrity, particularly in complex urban scenes. These quantitative

and qualitative results confirm the model’s robust segmentation capability on high-resolution imagery.

Table 2 Experimental results on the ISPRS vaihingen dataset

Model	Background	Imp. Surf	Building	Low. Veg	Tree	Car	mF1	mIoU
ABCNet	ResNet18	84.84	91.29	65.42	82.14	72.51	87.39	78.15
DeepLabV3 +	ResNet50	85.21	91.98	66.31	82.31	76.91	88.31	79.51
Unetformer	ResNet18	85.41	92.11	65.46	82.55	80.21	89.88	81.99
RS3Mamba	ResNet18	85.98	93.20	67.11	82.89	81.46	90.19	82.93
CM-Net	ResNet18	86.24	92.48	66.52	82.71	82.21	90.11	83.01
MCM-Net	ResNet18	86.37	93.12	67.23	83.11	82.71	90.73	83.65

Conclusions and Discussion

To address the complex challenges in remote sensing image semantic segmentation, a dual-branch segmentation model based on visual state space architecture was proposed in this study. The MCM-Net framework employs two parallel branches to extract local features (detail preservation) and global contextual representations (long-range dependencies) respectively. Experimental results on LoveDA and Vaihingen benchmarks demonstrate that MCM-Net outperforms state-of-the-art methods. This work provides a new paradigm for fusing local processing and global reasoning in remote sensing segmentation, with potential extensibility to other dense prediction tasks.

References

- [1] ZHANG D, WANG F, NING L, *et al.* Integrating SAM with feature interaction for remote sensing change detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1–11.
- [2] LI R, ZHENG S, DUAN C, *et al.* Land cover classification from remote sensing images based on multi-scale fully convolutional network [J].

Geo-spatial Information Science, 2022, 25(2): 278–294.

- [3] ZHAO L, ZHANG Y, SHI C, *et al.* APNet: A novel antiperturbation network for robust hyperspectral image classification against adversarial attacks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1–14.
- [4] HE D, SHI Q, LIU X, *et al.* Deep subpixel mapping based on semantic information modulated network for urban land use mapping [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(12): 10628–10646.
- [5] CHEN, K.; ZOU, Z.; SHI, Z. Building extraction from remote sensing images with Sparse Token Transformers [J]. *Remote Sens.* 2021, 13: 4441.
- [6] CAI J, TAO L, LI Y. CM-UNet + +: A multi-level information optimized network for urban water body extraction from high-resolution remote sensing imagery[J]. *Remote Sensing*, 2025, 17(6): 980.
- [7] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal[M]*. *Mach. Intell.* 2017.
- [8] CHEN LC, ZHU Y, PAPANDREOU G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [J]. in *Proc IEEE Eur Conf Comput Vis*, 2018: 801–818.

Editor: Yingzhi GUANG

Proofreader: Xinxiu ZHU

(Continued from page 37)

RCLA-PLE algorithm not only surpassed other multi-task models but also significantly outperformed all single-task models. Therefore, the RCLA-PLE model can effectively capture relationships within data while better utilizing potential information in different datasets to enhance predictive performance, establishing itself as an efficient multi-task learning algorithm.

References

- [1] ORGIAZZI A, BALLABIO C, PANAGOS P, *et al.* LUCAS Soil, the largest expandable soil dataset for Europe; A review[J]. *European Journal of Soil Science*, 2018, 69(1): 140–153.
- [2] BARNES RJ, DHANOA MS, LISTER SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra[J]. *Applied spectroscopy*, 1989, 43(5): 772–777.
- [3] TANG H, LIU J, ZHAO M, *et al.* Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations[C]//*Proceedings of the 14th ACM conference on recommender systems*. 2020: 269–278.
- [4] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning ap-

plied to document recognition[J]. *Proceedings of the IEEE*, 2002, 86(11): 2278–2324.

- [5] LI Y, ZHU Z, KONG D, *et al.* EA-LSTM: Evolutionary attention-based LSTM for time series prediction[J]. *Knowledge-Based Systems*, 2019, 181: 104785.
- [6] VANDENHENDE S, GEORGIOULIS S, VAN GANSBEKE W, *et al.* Multi-task learning for dense prediction tasks: A survey[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 44(7): 3614–3633.
- [7] PADARIAN J, MINASNY B, MCBRATNEY AB. Using deep learning to predict soil properties from regional spectral data[J]. *Geoderma Regional*, 2019, 16: e00198.
- [8] SINGH S, KASANA SS. Estimation of soil properties from the EU spectral library using long short-term memory networks[J]. *Geoderma Regional*, 2019, 18: e00233.
- [9] YANG J, WANG X, WANG R, *et al.* Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis – NIR spectroscopy[J]. *Geoderma*, 2020, 380: 114616.
- [10] TSAKIRIDIS NL, KERAMARIS KD, THEOCHARIS JB, *et al.* Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network[J]. *Geoderma*, 2020, 367: 114208.

Editor: Yingzhi GUANG

Proofreader: Xinxiu ZHU