Sensors and Materials, Vol. 36, No. 11 (2024) 5007–5017
MYU Tokyo

S & M 3846

LMetal-ResNet: A Lightweight Convolutional Neural Network Model for Soil Arsenic Concentration Estimation

Ai-Ping Wang,[1] An-Hong Tian,[2,1] and Cheng-Biao Fu[1*]

[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
[2]Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China

(Received June 25, 2024; accepted October 31, 2024)

Keywords: soil, visible and near-infrared spectroscopy, arsenic, convolutional neural network, lightweight

The application of hyperspectral remote sensing technology to soil environment monitoring is a cost-effective and time-saving methodology. Owing to the complex multidimensional and nonlinear relationships of hyperspectral data, traditional machine learning models are limited in their ability to deal with such complex multidimensional and nonlinear relationships. Deep learning models have been proven to effectively handle this complex multidimensional and nonlinear relational data. We took 183 soil samples from Mojiang Hani Autonomous County, Pu'er City, Yunnan Province and selected them as the research subjects, and the concentration of arsenic in the soil was predicted on the basis of the results of visible and near-infrared (Vis–NIR) spectroscopy, and proposed a lightweight convolutional neural network (CNN) model, LMetal-ResNet, aiming to predict soil arsenic concentration quickly and accurately. In addition, we also constructed two traditional machine learning models, partial least squares regression and support vector regression, and a CNN model, GoogleNet7, to predict arsenic concentration. In all the models, the data obtained by Vis–NIR spectroscopy with Savitzky–Golay convolution smoothing, min-max normalization preprocessing, and Pearson correlation coefficient feature band selection were used as input, and the output was soil arsenic concentration. Finally, the modeling accuracies of these four models were compared and analyzed. The experimental results showed that the modeling accuracies of the CNN models were higher than those of the traditional machine learning models, and LMetal-ResNet achieved the highest prediction accuracy using 68.59% of the parameters of GoogleNet7. LMetal-ResNet in the validation dataset had a root mean square error of 106.8862 mg/kg, a coefficient of determination of 0.8744, and a relative analytical error of 2.8221. In addition, we also analyzed the top 20 characteristic bands that contribute to LMetal-ResNet in predicting arsenic concentration using soil Vis–NIR spectroscopy data. This study provides scientific theoretical guidance and technical support for the prediction of concentration in soil using deep learning CNNs combined with hyperspectral remote sensing technology.

*Corresponding author: e-mail: fcb@kust.edu.cn
https://doi.org/10.18494/SAM5198

ISSN 0914-4935
https://myukk.org/

1. Introduction

Soil is the basis of human productive activities. Owing to the rapid development of urbanization and industrialization, the problem of soil pollution by heavy metals has become increasingly serious and poses serious health risks to plant growth and human health.[1,2] As a trace element, the heavy metal arsenic can promote the growth and development of animals and plants in appropriate amounts, but in excess, it can lead to serious toxic effects that affect the health and survival of plants and animals. Arsenic is currently recognized as number one on the list of substances that pose the most serious potential threat to human health, causing cancer and many types of organ toxicity.[3,4] The traditional method of detecting heavy metal concentrations in soil is to collect soil samples in the field and then bring the samples back to the laboratory for chemical analysis. This process is time-consuming and laborious. Therefore, the rapid and accurate detection of arsenic concentration in soil has become a focus of environmental scientists around the word. Hyperspectral remote sensing technology is now widely used in environmental research, precision agriculture, and other fields owing to its characteristics of high speed, low cost, and richness of band information. An increasing number of researchers are using this technology to quantitatively predict the physicochemical properties of soil.[5,6]

Most of the previous studies use traditional machine learning models combined with hyperspectral remote sensing technology to predict the various physical and chemical properties of soil. For example, Zhou *et al.* used first derivative, second derivative, inverse-log, continuum removal, and multivariate scattering correction as preprocessing methods for spectra. Partial least squares regression (PLSR), support vector machine (SVM), and random forest (RF) were used as prediction models to estimate the physical and chemical properties of manganese, lead, and zinc in soil. In the end, RF achieved the best modeling effect.[7] However, estimating heavy metal concentrations or other substances in soil using traditional machine learning models requires a great deal of specialized knowledge and experience in correctly selecting spectral preprocessing methods. There are also some difficulties in deeply or effectively revealing the complex nonlinear relationship between reflectance spectra and soil properties.[8]

In recent years, owing to deep learning excellent feature capturing abilities, some deep learning models have also been introduced in the field of hyperspectral remote sensing. CNN, as one of the classics among deep learning models, has achieved considerable success in many fields such as computer vision, natural language processing, and biomedical information.[9,10] Some studies have applied CNN models to soil hyperspectral inversion.[11] For example, Zhao *et al.* used a CNN model with an attention mechanism to model soil hyperspectral data to predict organic carbon and achieved good results.[12] These research results indicate the feasibility of applying CNN models in the prediction of soil components based on hyperspectral features, and they proved that CNN can effectively reveal the complex nonlinear relationship between soil reflectance spectra and soil properties. Despite the powerful learning capability of CNN models, in some resource-limited scenarios, such as edge devices with limited computational resources, the lightweighting of CNN models also needs to be considered. Therefore, in this study, a lightweight CNN model, LMetal-ResNet, was proposed, and the models PLSR, support vector regression (SVR), and GoogleNet7 were also constructed for the prediction of arsenic

concentration. All the models used visible and near-infrared spectral data processed by Savitzky–Golay (SG) convolution smoothing, min-max normalization, and Pearson correlation coefficients to predict soil arsenic concentration. In addition, the characteristic bands with the top 20 contribution degrees in predicting soil arsenic concentration by LMetal-ResNet are also listed.

In summary, the aims of this study are as follows: (1) The performance characteristics of convolutional neural network (CNN) and traditional machine learning algorithms in predicting soil arsenic concentration using Vis–NIR spectroscopy are compared. (2) A lightweight CNN model, LMetal-ResNet, was established to explore its potential to predict soil arsenic concentration. (3) The feature bands with the top 20 contribution degrees were analyzed when LMetal-ResNet was used to predict soil arsenic concentration.

2. Materials and Methods

2.1 Study area and data collection

The study area is located in Mojiang Hani Autonomous County, Pu'er City, Yunnan Province, China (101°08′–102°04′E, 22°51′–23°59′N). The research subject is the farmland soil in the area. According to the second soil census data of Mojiang County, lateritic red soil, latosol red soil, and red soil are the main types of soil in the area. There is also a gold mine in the study area. This gold mine has a long history since its development, and the mining process also resulted in some degree of contamination of the surrounding agricultural soils. Soil samples for this study were obtained from February 11 to 15, 2022, and a total of 183 soil samples were obtained by the grid method. Sampling requires clearing the surrounding materials such as stone particles, weeds, and other materials. The collection depth range of each sample is between 0 and 20 cm. The collected soil samples were then taken back to the laboratory for soil Vis–NIR spectroscopy and arsenic concentration determination. Vis–NIR spectroscopy was conducted in a dark room using an American ASD FieldSpec3 spectrometer (band range: 350–2500 nm), and the arsenic concentration was determined by chemical analysis.

2.2 Spectral data preprocessing and feature band selection

For subsequent experimental modeling and improving the predictive performance of the model, the preprocessing of the spectral data and the selection of the feature bands are also required. We first used the five-point two-order SG convolution smoothing algorithm to denoise the spectral data. Then, to eliminate the impact of scale differences on the final experimental results and accelerate the convergence speed of the model, the spectral data and arsenic concentration were then subjected to min-max normalization processing. The min-max normalization calculation formula is

$$y = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{1}$$

where x and y represent the data before and after normalization, respectively. x_{min} and x_{max} represent the minimum and maximum absorbance values of the samples at the same wavelength, or the minimum and maximum values of arsenic in the sample, respectively. The feature band selection used the Pearson correlation coefficient feature selection method. We selected 947 spectral bands with the absolute values of correlation coefficients greater than 0.6 as feature bands of the input model.

2.3 Modeling methods

2.3.1 PLSR and SVR

PLSR is a statistical modeling technique that is mainly used to deal with the relationship between multivariate data. SVR is a regression algorithm based on SVM, which uses the maximum interval idea and kernel trick of SVM to deal with regression problems. The main modeling process for the PLSR and SVR models in this study was to input the preprocessed Vis–NIR spectral data into the two prediction models. Then, parameter optimization was performed using the grid search method, and the final model output was the arsenic content.

2.3.2 CNN

CNN is a deep learning model that is commonly used in visual processing and automatically learns feature representations mainly through convolutional and pooling layers. The network structure diagrams of the two CNN models used in this study, GoogleNet7 and LMetal-ResNet, are shown in Figs. 1 and 2, respectively. The modeling process is mainly divided into the following: (1) The training set spectral data with dimensions (122947) after preprocessing and feature band selection are input to the input layer, where 122 denotes the number of training datasets and 947 denotes the number of spectral features. (2) Then, the training dataset passes through the convolution layer, pooling layer, and other network layers, and outputs a predicted value in the range of 0–1 at the output layer, and then, reverse min-max normalization is carried

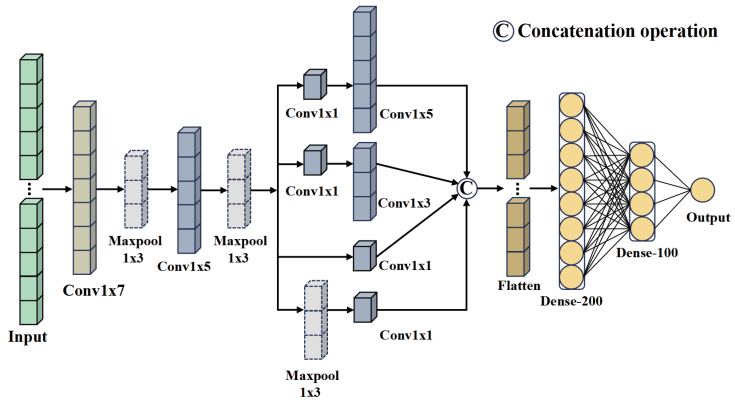


Fig. 1. (Color online) GoogleNet7 structure diagram.

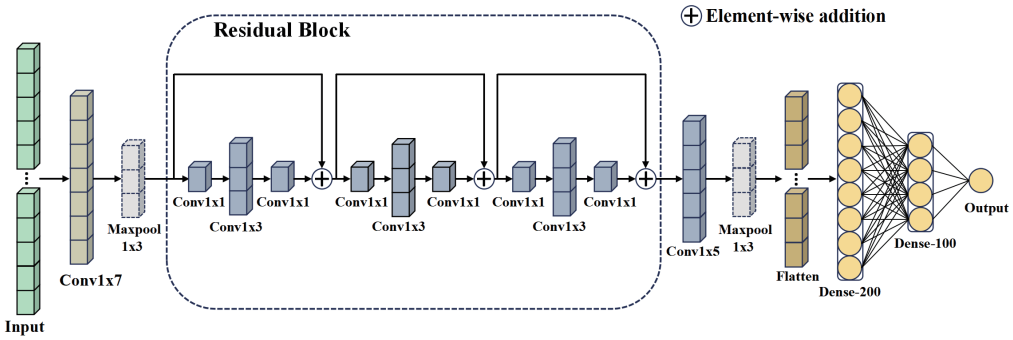


Fig. 2. (Color online) LMetal-ResNet structure diagram.

out to obtain the predicted value of arsenic. During the entire training process, the model updates the network layer weights through continuous iterative training to minimize the loss function value and save the best network weight file. (3) After the model training was completed, the trained model was evaluated using the validation dataset in combination with evaluation metrics.

To make the CNN models comparable, each model needs to fix the activation function and the values of certain hyperparameters. The specific settings are as follows: 1D convolutional kernels are used for all convolutional layers, the learning rate is 0.001, the optimizer is Nadam, the activation function for the output layer is Sigmoid, and the activation function for the other network layers is Tanh. The number of training Epochs is 2000, and the inactivation rate of the Dropout layer is 50%. To avoid model overfitting, an early stopping strategy is used, and the number of early stops is 200.

2.4 Evaluation criteria for the model

We used *RMSE*, R^2, and *RPD* as evaluation indicators of the prediction model. *RMSE* is used to characterize the difference between predicted and measured values, with a smaller *RMSE* indicating a better performance of the model. When R^2 is closer to 1, the prediction accuracy of the model is considered high. *RPD* is used to assess the predictive ability of the model. When $RPD < 1.0$, the model is considered unable to carry out the prediction task. When $1.0 < RPD < 1.4$, the model has poor predictive performance. When $1.4 < RPD < 1.8$, the model has average predictive performance. When $2.0 < RPD < 2.5$, the model has very good predictive performance, and when $RPD > 2.5$, the model has excellent predictive performance. [13,14]

3. Results and Discussion

3.1 Statistical description of arsenic concentration in soil

The statistical description of arsenic concentration in the 183 experimental soil samples used in this study is shown in Table 1. Table 1 shows that the arsenic concentrations of these 183 soil samples varied from 5.79 to 1780 mg/kg with a coefficient of variation of 128.8805%. This shows that there is considerable variability in the geographical spatial distribution of soil arsenic concentration in this region.[15] We used the concentration ranking scheme to divide the dataset of these 183 soil samples in a 2:1 ratio. The soil samples were sorted in descending order according to arsenic concentration, and then every three samples were grouped into one set. The first two in a group were divided into the training dataset and the remaining one was the validation dataset. Finally, the number of training datasets was 122 and the number of validation datasets was 61. The coefficient of variation of the training datasets was 131.0852%, and the coefficient of variation of the validation datasets was 123.1885%. The coefficient of variation between the two was not much different, indicating that the distribution difference between the divided training and validation datasets was small. As a result, the large geographical spatial variability of arsenic concentration in the study area will not affect the final evaluation results.

3.2 Comparison of modeling accuracies of different prediction models

The prediction accuracies of the four models, PLSR, SVR, GoogleNet7, and LMetal-ResNet, are shown in Table 2. In the validation dataset, the *RPD*s of the four prediction models all reached above 1.8, and the prediction accuracy of the CNN model was greater than that of the traditional machine learning model. Among all the models, LMetal-ResNet had the highest prediction accuracy, with $RMSE$ = 91.1363 mg/kg, R^2 = 0.9305, and RPD = 3.7932 in the training dataset. In the validation dataset, $RMSE$ = 106.8862 mg/kg, R^2 = 0.8744, and RPD = 2.8221. The RPD of LMetal-ResNet in the validation dataset was greater than 2.5, indicating its excellent

Table 1
Statistical description of soil arsenic.

Attributes	Sample datasets	Num	Max	Min	Mean	SD	CV (%)
Arsenic (mg/kg)	Entire	183	1780	5.79	257.4367	331.7858	128.8805
	Training	122	1780	5.98	263.7216	345.6999	131.0852
	Validation	61	1372	5.79	244.8670	301.6480	123.1885

Table 2
Accuracies of four models in predicting arsenic.

Model	Training dataset			Validation dataset		
	RMSE	R^2	*RPD*	*RMSE*	R^2	*RPD*
PLSR	122.4128	0.8746	2.8241	166.6468	0.6948	1.8101
SVR	146.6997	0.8199	2.3565	152.9526	0.7429	1.9722
GoogleNet7	96.7499	0.9217	3.5731	107.9214	0.8720	2.7951
LMetal-ResNet	91.1363	0.9305	3.7932	106.8862	0.8744	2.8221

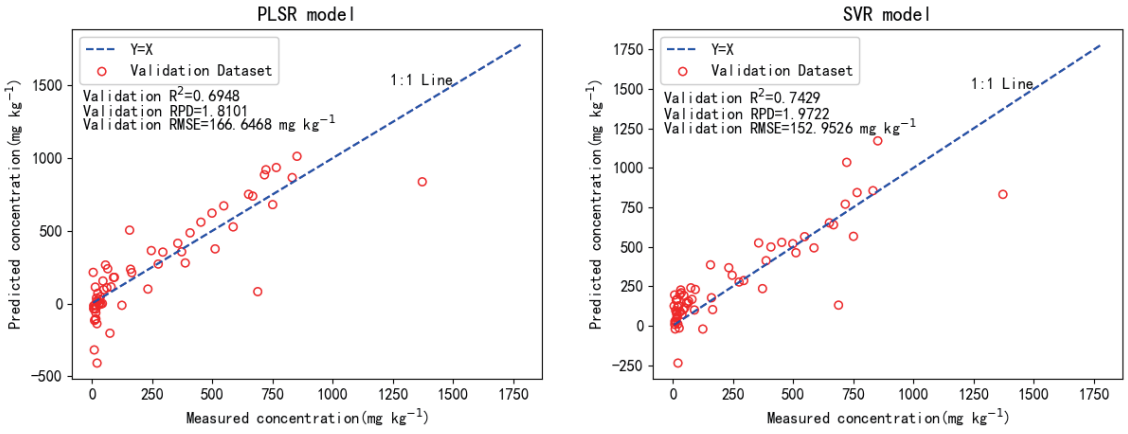
predictive ability. The model with the second highest prediction accuracy was GoogleNet7, which had *RMSE* = 107.9214 mg/kg, R^2 = 0.8720, and *RPD* = 2.7951 in the validation dataset. Its predictive ability was also excellent. For traditional machine learning models, SVR had the highest prediction accuracy, and PLSR had the lowest prediction accuracy. In the validation dataset, *RMSE* = 152.9526 mg/kg, R^2 = 0.7429, and *RPD* = 1.9722 for SVR, and *RMSE* = 166.6468 mg/kg, R^2 = 0.6948, and *RPD* = 1.8101 for PLSR. The accuracies of both models are average. LMetal-ResNet has the highest prediction accuracy compared with PLSR, which has the lowest prediction accuracy; *RMSE* decreased by 59.7606 mg/kg, R^2 increased by 0.1796, and *RPD* increased by 1.012.

3.3 Discussion

Previous studies using Vis–NIR spectroscopy to predict soil heavy metals mostly used traditional machine learning models, while there were fewer studies using CNN models in deep learning for prediction. Deep learning CNN has significant advantages over traditional machine learning models when dealing with complex nonlinear relationships and high-dimensional Vis–NIR spectroscopy data.[16] However, deep learning models usually have a large number of parameters and complex network architecture. This results in a large amount of calculation and memory consumption, so that its deployment on edge devices faces huge challenges.[17]

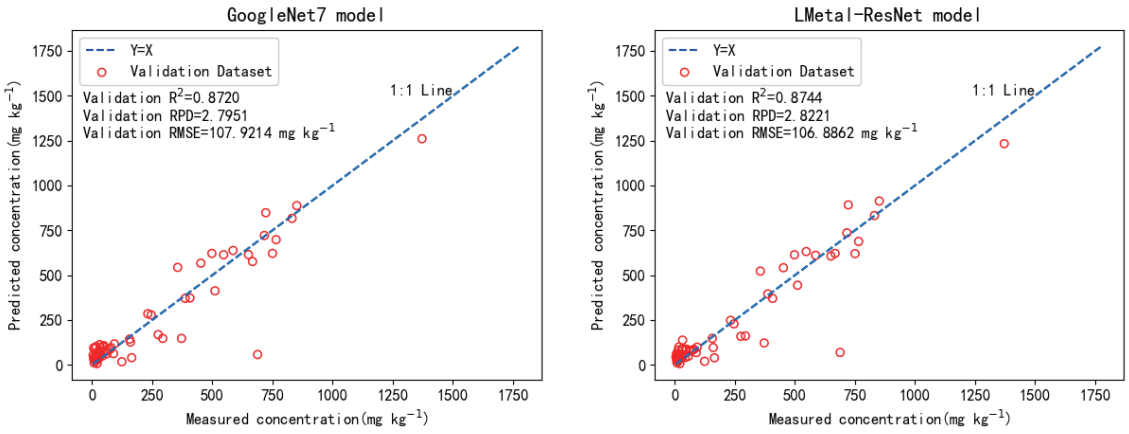
Therefore, in this study, we propose a lightweight CNN model, LMetal-ResNet, to predict arsenic concentration in soil. The experimental results in Table 1 show that the CNN models are better than the traditional machine learning models, and LMetal-ResNet achieved the highest prediction accuracy (in the validation dataset, *RPD* = 2.8221 and R^2 = 0.8744). This is then analyzed in conjunction with the scatterplot of the four models in the validation dataset, as shown in Fig. 3.

On the scatter plot, the predicted points of the CNN model are closer to the 1:1 fitting line than those of the traditional machine learning model. This is consistent with the experimental results of Pyo *et al.* using a CNN model with an autoencoder combined with visible and near-



(a) (b)

Fig. 3. (Color online) Scatter plots of (a) PLSR, (b) SVR, (c) GoogleNet7, and (d) LMetal-ResNet on the validation dataset.



(c) (d)

Fig. 3. (Color online) (continued) Scatter plots of (a) PLSR, (b) SVR, (c) GoogleNet7, and (d) LMetal-ResNet on the validation dataset.

infrared spectroscopy for heavy metals such as arsenic in soil (R^2 = 0.86).[18] The prediction accuracy of LMetal-ResNet is even higher. Additionally, the traditional machine learning models, PLSR and SVR, both exhibited instances of predicted points with negative values. Then, we analyzed the scatter plots of the two CNN models GoogleNet7 and LMetal-ResNet. We found that the fitting lines of the two scatter plots are roughly the same. Given that the prediction accuracies of the two are roughly the same, we further analyzed the parameter quantities of the models. The parameter quantity statistics are shown in Table 3. Table 3 shows that the number of parameters in LMetal-ResNet is reduced by 31.41% relative to that in GoogleNet7. Moreover, it maintains almost the same prediction accuracy as GoogleNet7 or even better, even when the number of parameters is reduced. We analyzed the reason and found that it is mainly because the introduction of the residual learning module allows LMetal-ResNet to avoid the loss of feature information when stacking multiple network layers. A shortcut connection between layers enables the network to be effectively trained and optimized even as the network depth increases, and problems such as gradient vanishing and gradient exploding are alleviated.[19] In the residual learning module of LMetal-ResNet, a large number of 1 × 1 convolution kernels are used. According to the calculation [Eq. (2)] of the convolution layer parameters, using a 1 × 1 convolution kernel can effectively reduce the number of parameters of the model.

$$p = c_{out} * \left(k_w * k_h * c_{in}\right) + b \tag{2}$$

Here, p represents the number of parameters, c_{in} and c_{out} represent the numbers of channels of the feature map input and output, respectively. k_w and k_h represent the width and height of the convolution kernel, respectively, and b represents the number of parameters calculated for the bias term ($b = c_{out}$).

We also used the deep learning interpreter DeepExplainer in the SHAP library to analyze the contribution of the characteristic bands when LMetal-ResNet predicts arsenic concentration in the soil. This interpreter is inspired by game theory and uses the model's internal information to

Table 3
Statistics on the parameters of the convolutional and dense layers of GoogleNet7 and LMetal-ResNet.

Model	Convolutional layer	Dense layer	Total parameters
GoogleNet7	936	434401	435337
LMetal-ResNet	2214	296401	298615

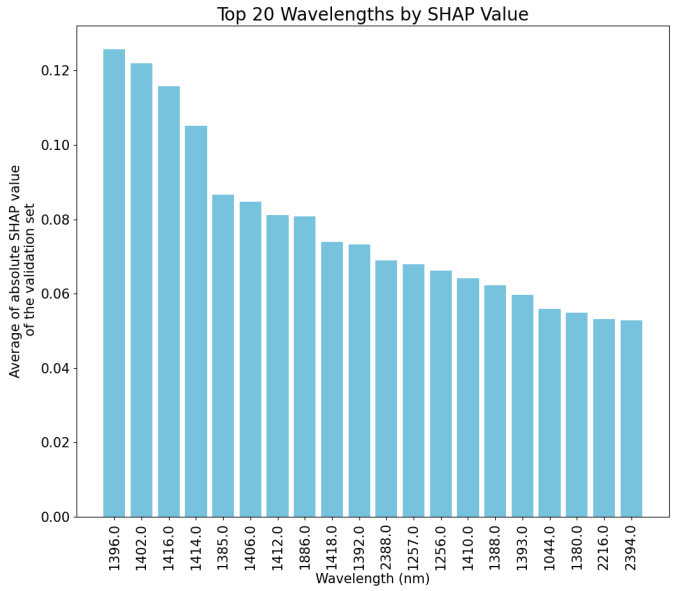


Fig. 4. (Color online) Feature bands with the top 20 contribution degrees.

calculate the SHAP value.[20] If the SHAP average absolute value of a certain feature is high, the feature is considered to contribute more to the model's prediction task. We used the validation set to conduct a feature band contribution test experiment on LMetal-ResNet, and the feature bands with the top 20 contribution degrees were extracted. The results are shown in Fig. 4.

4. Conclusions

In this paper, a lightweight CNN model, LMetal-ResNet, was proposed. The model uses the results of the preprocessed Vis–NIR spectroscopy of soil as input to the model to predict arsenic concentration. We compared LMetal-ResNet with two traditional machine learning models, PLSR and SVR, as well as the CNN model GoogleNet7 in deep learning. The experimental results showed that the deep learning CNN models exhibit higher prediction accuracies than the traditional machine learning models. In particular, LMetal-ResNet outperformed the prediction accuracy of GoogleNet7 when using 31.41% fewer parameters than GoogleNet7 ($RMSE$ = 106.8862 mg/kg, R^2 = 0.8744, and RPD = 2.8221). The results of this study can provide theoretical and technical guidance for predicting arsenic concentration using soil Vis–NIR spectroscopy.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (42067029,42361007), the Reserve Talents Project for Young and Middle-aged Academic and Technical Leaders in Yunnan Province (202205AC160005), and the Yunnan Province "Xing Dian Ying Talent Support Programme" Young Talents Project (KKXX202303001).

References

1 T. Shi, H. Liu, Y. Chen, J. Wang, and G. Wu: J. Hazard. Mater. **308** (2016) 243. https://doi.org/10.1016/j.jhazmat.2016.01.022
2 X. Cui, Y. Geng, R. Sun, M. Xie, X. Feng, X. Li, and Z. Cui: J. Cleaner Prod. **295** (2021) 126504. https://doi.org/10.1016/j.jclepro.2021.126504
3 M. U. Rehman, R. Khan, A. Khan, W. Qamar, A. Arafah, A. Ahmad, A. Ahmad, R. Akhter, J. Rinklebe, and P. Ahmad: J. Hazard. Mater. **417** (2021) 126050. https://doi.org/10.1016/j.jhazmat.2021.126050
4 M. K. Yadav, D. Saidulu, A. K. Gupta, P. S. Ghosal, and A. Mukherjee; J. Environ. Chem. Eng. **9** (2021) 105203. https://doi.org/10.1016/j.jece.2021.105203
5 F. Guo, Y. Wang, D. Lin, and Z. Xu: IEEE Geosci. Remote Sens. Lett. **20** (2023) 1. https://doi.org/10.1109/LGRS.2023.3330854
6 W. Sun, S. Liu, X. Zhang, and Y. Li: Geoderma **409** (2022) 115653. https://doi.org/10.1016/j.geoderma.2021.115653
7 W. Zhou, H. Yang, L. Xie, H. Li, L. Huang, Y. Zhao, and T. Yue: CATENA **202** (2021) 105222. https://doi.org/10.1016/j.catena.2021.105222
8 J. Padarian, B. Minasny, and A. B. McBratney: Geoderma **425** (2022) 116063. https://doi.org/10.1016/j.geoderma.2022.116063
9 A. Singh, and L. Bruzzone: IEEE Geosci. Remote Sens. Lett. **19** (2022) 1. https://doi.org/10.1109/LGRS.2022.3211861
10 S. Behera, D. P. Dogra, M. K. Bandyopadhyay, and P. P. Roy: IEEE Trans. Cybern. **53** (2023) 3428. https://doi.org/10.1109/TCYB.2021.3126434
11 L. Zhong, X. Guo, Z. Xu, and M. Ding: Geoderma **402** (2021) 115366. https://doi.org/10.1016/j.geoderma.2021.115366
12 W. Zhao, Z. Wu, Z. Yin, and D. Li: IEEE Geosci. Remote Sens. Lett. **19** (2022) 1. https://doi.org/10.1109/LGRS.2022.3201266
13 Y. Mao, H. Li, Y. Wang, H. Wang, J. Shen, Y. Xu, S. Ding, H. Wang, Z. Ding, and K. Fan: Comput. Electron. Agric. **213** (2023) 108176. https://doi.org/10.1016/j.compag.2023.108176
14 Z. M. Yaseen: Chemosphere **277** (2021) 130126. https://doi.org/10.1016/j.chemosphere.2021.130126
15 S. Moradpour, M. Entezari, S. Ayoubi, A. Karimi, and S. Naimi: J. Hazard. Mater. **455** (2023) 131609. https://doi.org/10.1016/j.jhazmat.2023.131609
16 Y. Wang, S. Chen, Y. Hong, B. Hu, J. Peng, and Z. Shi: Comput. Electron. Agric. **212** (2023) 108067. https://doi.org/10.1016/j.compag.2023.108067
17 P. Liu, J. Du, and C.-M. Vong: Neural Networks **164** (2023) 124. https://doi.org/10.1016/j.neunet.2023.04.023
18 J. Pyo, S. M. Hong, Y. S. Kwon, M. S. Kim, and K. H. Cho: Sci. Total Environ. **741** (2020) 140162. https://doi.org/10.1016/j.scitotenv.2020.140162
19 Y. Wang, A. Abliz, H. Ma, L. Liu, A. Kurban, U. Halik, M. Pietikainen, and W. Wang: IEEE Trans. Geosci. Remote Sens. **60** (2022) 1. https://doi.org/10.1109/TGRS.2022.3190310
20 S. M. Lundberg and S.-I. Lee: Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS'17) 4768

About the Authors

Aiping Wang received his bachelor's degree from the School of Software at East China University of Technology in 2023. He is currently pursuing a master's degree in the School of Information Engineering and Automation, Kunming University of Science and Technology. His main research interests include hyperspectral remote sensing and artificial intelligence. (wap@stu.kust.edu.cn)

Anhong Tian received her B.S. and M.S. degrees from Chongqing University of Posts and Telecommunications, China, in 2007 and 2010, respectively. She received her Ph.D. degree from Kunming University of Science and Technology, China, in 2020. Since 2023, she has been a professor at Kunming University of Science and Technology. Her research interests are in hyperspectral remote sensing, nonlinear systems, and artificial intelligence. (tah@kust.edu.cn)

Chengbiao Fu received his B.S. and M.S. degrees from Chongqing University of Posts and Telecommunications, China, in 2005 and 2009, respectively. He received his Ph.D. degree from Kunming University of Science and Technology, China, in 2021. Since 2023, he has been an associate professor at Kunming University of Science and Technology. His research interests are in information identification and processing, hyperspectral remote sensing and machine learning. (fcb@kust.edu.cn)